Date: 1/30/04 Express Mail Label No. EV 052030 79345

Inventor: Anna Helgadottir

Attorney's Docket No.: 2345.2048-003

SUSCEPTIBILITY GENE FOR MYOCARDIAL INFARCTION AND STROKE

RELATED APPLICATION

This application is a continuation-in-part of International Application No. PCT/US03/32805, which designated the United States and was filed on October 16, 2003, published in English, which claims the benefit of U.S. Provisional Application No. 60/419,432, filed on October 17, 2002.

The entire teachings of the above applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Myocardial infarction (MI) is one of the most common diagnoses in hospitalized patients in industrialized countries. Myocardial Infarction generally occurs when there is an abrupt decrease in coronary blood flow following a thrombotic occlusion of a coronary artery previously narrowed by atherosclerosis. Infarction occurs when a coronary artery thrombus develops rapidly at a site a vascular injury, which is produced or facilitated by factors such as cigarette smoking, hypertension and lipid accumulation. In most cases, infarction occurs when an atherosclerotic plaque fissures, ruptures or ulcerates and when conditions favor thrombogenesis. In rare cases, infarction may be due to coronary artery occlusion caused by coronary emboli, congenital abnormalities, coronary spasm, and a wide variety of systemic, particularly inflammatory diseases.

Although classical risk factors such as smoking, hyperlipidemia, hypertension, and diabetes are associated with many cases of coronary heart disease (CHD) and MI, many patients do not have involvement of these risk factors. In fact, many patients who exhibit one or more of these risk factors do not develop MI. Family history has long been recognized as one of the major risk factors. Although some of the familial clustering of MI reflects the genetic contribution to the other conventional risk factors, a large number of studies have suggested that there are significant genetic susceptibility factors, beyond those of the known risk factors (Friedlander Y, *et al., Br Heart J.* 1985; 53:382-7, Shea S. *et al., J. Am. Coll. Cardiol.* 1984; 4:793-801, and Hopkins P.N., *et al., Am. J. Cardiol.* 1988; 62:703-7). Major genetic susceptibility factors have not yet been identified.

SUMMARY OF THE INVENTION

As described herein, a locus on chromosome 13q12 has been identified as playing a major role in Myocardial Infarction (MI). The locus, herein after referred to as the MI locus, comprises nucleic acid that encodes 5-lipoxygenase activating protein (ALOX5AP or FLAP), herein after referred to as FLAP. The gene has also been shown to play a role in stroke.

The present invention relates to isolated nucleic acid molecules comprising a portion or the entire human FLAP nucleic acid or a variant thereof. In one embodiment, the nucleic acid molecule has at least one polymorphism that is correlated with the incidence of myocardial infarction or stroke. Identification of nucleic acids and polymorphisms in this locus can pave the way for a better understanding of the disease process, which in turn can lead to improved diagnostic and therapeutic methods.

The invention further pertains to methods of diagnosing a susceptibility to myocardial infarction or stroke, comprising detecting an alteration in the expression or composition of a polypeptide encoded by a FLAP nucleic acid in a test sample, in comparison with the expression or composition of a polypeptide encoded by FLAP in a control sample, wherein the presence of an alteration in expression or composition

of the polypeptide in the test sample is indicative of a susceptibility to myocardial infarction or stroke.

The invention also relates to an isolated nucleic acid molecule comprising a FLAP nucleic acid, wherein the FLAP nucleic acid has a nucleic acid sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or the complement of SEQ ID NO: 1 or SEQ ID NO: 3, wherein the nucleic acid molecule comprises a polymorphism as indicated in Table 3.

In another embodiment, the invention relates to an isolated nucleic acid molecule having a polymorphism as indicated in Table 3, which hybridizes under high stringency conditions to a nucleic acid sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or the complement of SEQ ID NO: 1 or SEQ ID NO: 3.

In yet another embodiment, a method for assaying for the presence of a first nucleic acid molecule in a sample is described, comprising contacting said sample with a second nucleic acid molecule, where the second nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 or SEQ ID NO: 3, and hybridizes to the first nucleic acid under high stringency conditions.

The invention also relates to a vector comprising an isolated nucleic acid molecule of the invention operably linked to a regulatory sequence, as well as to a recombinant host cell comprising the vector. The invention also provides a method for preparing a polypeptide encoded by an isolated nucleic acid molecule comprising culturing the recombinant host cell under conditions suitable for expression of said nucleic acid molecule.

Also contemplated by the invention is a method of assaying a sample for the presence of a polypeptide encoded by an isolated nucleic acid molecule of the invention, comprising contacting the sample with an antibody that specifically binds to the polypeptide.

The invention further provides a method of identifying an agent that alters expression of a FLAP nucleic acid, comprising: contacting a solution containing a nucleic acid comprising the promoter region of the FLAP nucleic acid operably linked to a reporter gene with an agent to be tested; assessing the level of expression of the reporter gene; and comparing the level of expression with a level of expression of the

reporter gene in the absence of the agent; wherein if the level of expression of the reporter gene in the presence of the agent differs, by an amount that is statistically significant, from the level of expression in the absence of the agent, then the agent is an agent that alters expression of the FLAP nucleic acid. An agent identified by this method is also contemplated.

The invention additionally comprises a method of identifying an agent that alters expression of a FLAP nucleic acid, in which a solution containing a nucleic acid described herein or a derivative or fragment thereof is contacted with an agent to be tested, and expression of the nucleic acid, derivative or fragment in the presence of the agent is assessed and compared with expression of the nucleic acid, derivative or fragment in the absence of the agent. If expression of the nucleic acid, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant, from the expression in the absence of the agent, then the agent is an agent that alters expression of the FLAP nucleic acid. In certain embodiments, the expression of the nucleic acid, derivative or fragment in the presence of the agent comprises expression of one or more splicing variant(s) that differ in kind or in quantity from the expression of one or more splicing variant(s) the absence of the agent. Agents identified by this method are also contemplated. Representative agents include antisense nucleic acid to a FLAP nucleic acid; a FLAP polypeptide; a FLAP nucleic acid receptor; a FLAP nucleic acid binding agent; a peptidomimetic; a fusion protein; a prodrug thereof; an antibody; and a ribozyme. A method of altering expression of a FLAP nucleic acid comprising contacting a cell containing a FLAP nucleic acid with such an agent is also contemplated.

The invention further pertains to a method of identifying a polypeptide which interacts with a FLAP polypeptide, employing a yeast two-hybrid system that uses a first vector which comprises a nucleic acid encoding a DNA binding domain and a FLAP polypeptide, splicing variant, or a fragment or derivative thereof, and a second vector which comprises a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide. If transcriptional activation occurs in the yeast two-hybrid system, the test polypeptide is a polypeptide which interacts with a FLAP polypeptide.

A transgenic animal comprising a nucleic acid of the invention such as an exogenous FLAP nucleic acid or a nucleic acid encoding a FLAP polypeptide is also contemplated.

In yet another embodiment, the invention relates to a method for assaying a sample for the presence of a FLAP nucleic acid, by contacting the sample with a nucleic acid comprising a contiguous nucleic acid sequence which is at least partially complementary to a part of the sequence of said FLAP nucleic acid, under conditions appropriate for hybridization, and assessing whether hybridization has occurred between a FLAP nucleic acid and said nucleic acid, wherein if hybridization has occurred, a FLAP nucleic acid is present in the nucleic acid. In certain embodiments, the contiguous nucleic acid sequence is completely complementary to a part of the sequence of said FLAP nucleic acid and in other embodiments; amplification is of at least part of said FLAP nucleic acid.

In certain embodiments, the contiguous nucleic acid sequence is 100 or fewer nucleotides in length and is either: a) at least 80% identical to a contiguous sequence of nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3; b) at least 80% identical to the complement of a contiguous sequence of nucleotides in of SEQ ID NO: 1 or SEQ ID NO: 3; or c) capable of selectively hybridizing to said FLAP nucleic acid.

The invention also pertains to a reagent for assaying a sample for the presence of a FLAP nucleic acid, the reagent comprising a nucleic acid comprising a contiguous nucleic acid sequence which is at least partially complementary to a part of the nucleic acid sequence of said FLAP nucleic acid. The reagent can comprise a contiguous nucleotide sequence which is completely complementary to a part of the nucleic acid sequence of said FLAP nucleic acid. A reagent kit for assaying a sample for the presence of a FLAP nucleic acid is also described, including (*e.g.*, in separate containers), one or more labeled nucleic acids comprising a contiguous nucleic acid sequence which is at least partially complementary to a part of the nucleic acid sequence of said FLAP nucleic acid; and reagents for detection of said label. The labeled nucleic acid can comprise a contiguous nucleotide sequence which is completely complementary to a part of the nucleic acid sequence of said FLAP nucleic acid. Also described herein is a reagent kit for assaying a sample for the

presence of a FLAP nucleic acid, comprising one or more nucleic acids comprising a contiguous nucleic acid sequence which is at least partially complementary to a part of the nucleic acid sequence of said FLAP nucleic acid, and which is capable of acting as a primer for said FLAP nucleic acid when maintained under conditions for primer extension.

The invention also provides for the use of a nucleic acid for assaying a sample for the presence of a FLAP nucleic acid, in which the nucleic acid is 100 or fewer nucleotides in length and is either: at least 80% identical to a contiguous sequence of nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3; at least 80% identical to the complement of a contiguous sequence of nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3; or capable of selectively hybridizing to said FLAP nucleic acid.

In yet another embodiment, the use of a first nucleic acid for assaying a sample for the presence of a FLAP nucleic acid that has at least one nucleotide difference from the first nucleic acid is described, in which the first nucleic acid is 100 or fewer nucleotides in length and which is either: at least 80% identical to a contiguous sequence of nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 or one of the sequences shown in Table 3; at least 80% identical to the complement of a contiguous sequence of nucleotides of SEQ ID NO: 1 or SEQ ID NO: 3 one of the sequences shown in Table 3; or capable of selectively hybridizing to said FLAP nucleic acid.

The invention also relates to a method of diagnosing a susceptibility to myocardial infarction or stroke in an individual, comprising determining the presence or absence in the individual of certain "haplotypes" (combinations of genetic markers); the presence of the haplotype is diagnostic of susceptibility to myocardial infarction or stroke. In one embodiment, a haplotype associated with a susceptibility to myocardial infarction or stroke comprises markers DG00AAFIU, SG13S25, DG00AAJFF, DG00AAHII, SG13S32 and SG13S35 at the 13q12 locus. In one particular embodiment, the presence of the alleles T, G, G, G, A and G at DG00AAFIU, SG13S25, DG00AAJFF, DG00AAHII, SG13S32 and SG13S35, respectively (the B6 haplotype), is diagnostic of susceptibility to myocardial infarction or stroke. In another embodiment, a haplotype associated with a susceptibility to myocardial infarction or stroke comprises markers DG00AAFIU,

SG13S25, DG00AAHII, SG13S30 and SG13S42 at the 13q12 locus. In one particular embodiment, the presence of the alleles T, G, G, G and A at DG00AAFIU, SG13S25, DG00AAHII, SG13S30 and SG13S42, respectively (the B5 haplotype), is diagnostic of susceptibility to myocardial infarction or stroke. In a third embodiment, a haplotype associated with a susceptibility to myocardial infarction or stroke comprises markers SG13S25, DG00AAHII, SG13S30 and SG13S42 at the 13q12 locus. In one particular embodiment, the presence of the alleles G, G, G and A at SG13S25, DG00AAHII, SG13S30 and SG13S42, respectively (the B4 haplotype), is diagnostic of susceptibility to myocardial infarction or stroke. In a fourth embodiment, a haplotype associated with a susceptibility to myocardial infarction or stroke comprises markers DG00AAFIU, SG13S25, DG00AAHID, B_SNP_310657 and SG13S32 at the 13q12 locus. In one particular embodiment, the presence of the alleles T, G, T, G and A at DG00AAFIU, SG13S25, DG00AAHID, B_SNP_310657 and SG13S32, respectively (the A5 haplotype), is diagnostic of susceptibility to myocardial infarction or stroke. In a fifth embodiment, a haplotype associated with a susceptibility to myocardial infarction or stroke comprises markers SG13S25, DG00AAHID, B_SNP_310657 and SG13S32 at the 13q12 locus. In one particular embodiment, the presence of the alleles G, T, G and A at SG13S25, DG00AAHID, B_SNP_310657 and SG13S32, respectively (the A4 haplotype), is diagnostic of susceptibility to myocardial infarction or stroke. The presence or absence of the haplotype can be determined by various methods, including, for example, using enzymatic amplification, restriction fragment length polymorphism analysis, sequence analysis or electrophoretic analysis of nucleic acid from the individual.

A further embodiment of the invention is a method for identification of susceptibility to myocardial infarction or stroke, by identifying haplotypes and SNPs that can be used to identify individuals at risk of developing MI or stroke. The haplotypes can comprise, for example, at least one of the polymorphisms as indicated in Table 3, or as shown in the haplotypes in Table 4, Table 5, Table 7, and/or Table 13. In certain additional embodiments, the haplotype can be one of the haplotypes in Table 4, Table 5, Table 7, or Table 13; in other embodiments, the haplotype can be haplotypes B4, B5, B6, A4, A5 or Hap B.

A method for the diagnosis and identification of susceptibility to myocardial infarction in an individual is also described, comprising: screening for an at-risk haplotype in the FLAP nucleic acid that is more frequently present in an individual susceptible to myocardial infarction compared to an individual who is not susceptible to myocardial infarction wherein the at-risk haplotype increases the risk significantly. In certain embodiments, the significant increase is at least about 20%, and in other embodiments, the significant increase is identified as an odds ratio of at least about 1.2.

An additional embodiment comprises methods for the diagnosis of increased risk of susceptibility to myocardial infarction or stroke in an individual, by screening for an at-risk haplotype in the FLAP nucleic acid that is more frequently present in an individual susceptible to myocardial infarction or stroke (affected), compared to the frequency of its presence in a healthy individual (control). The presence of the at-risk haplotype is indicative of a susceptibility to myocardial infarction or stroke. In one embodiment, the at-risk haplotype has a p value < 0.05. In certain other embodiments, the screening for the presence of an at-risk haplotype comprises screening for an at-risk haplotype within or near FLAP that significantly correlates with a haplotype such as a halotype shown in Table 4; a haplotype shown in Table 5; a haplotype shown in Table 13; haplotype B4; haplotype B5; haplotype B6; haplotype A4; haplotype A5; or haplotype HapB. In other embodiments, screening for the presence of an at-risk haplotype comprises screening for an at-risk haplotype within or near FLAP that significantly correlates with susceptibility to myocardial infarction or stroke.

A further embodiment comprises methods of diagnosing FLAP-associated myocardial infarction or stroke in an individual who has had a myocardial infarction and/or a stroke, by detecting a polymorphism in a FLAP nucleic acid, or an alteration in the expression or composition of a polypeptide encoded by a flap nucleic acid, wherein the presence of the polymorphism in the nucleic acid or the alteration in expression or composition is indicative of FLAP-associated myocardial infarction or stroke. Additional embodiments of the invention include methods for identification of FLAP-associated myocardial infarction or stroke, by identifying haplotypes and SNPs

associated with MI or stroke. The haplotypes can comprise, for example, at least one of the polymorphisms as indicated in Table 3, or as shown in the haplotypes in Table 4, Table 5, Table 7, and/or Table 13. In certain additional embodiments, the haplotype can be one of the haplotypes in Table 4, Table 5, Table 7, or Table 13; in other embodiments, the haplotype can be haplotypes B4, B5, B6, A4, A5 or Hap B.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention. The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1 shows the multipoint non-parametric LOD scores for a framework marker map on chromosome 13. A LOD score suggestive of linkage of 2.5 was found at marker D13S289. The marker map for chromosome 13 that was used in the linkage analysis is shown in Table 1.

FIG. 2 shows LOD score results for the families after adding 14 markers to the candidate region. The inclusion of additional microsatellite markers increased the information on sharing by descent from 0.7 to 0.8, around the markers that gave the highest LOD scores. The marker map used in the second step of linkage anaysis is shown in Table 2.

FIG. 3.1 shows the results from a haplotype association analysis using 4 and 5 microsatellite markers. The *p*-value of the association is plotted on the y-axis and position of markers on the x-axis. Only haplotypes that show association with a *p*-value $< 10^{-5}$ are shown in the figure. The most significant microsatellite marker haplotype association is found using markers DG13S1103, DG13S166, DG13S1287, DG13S1061 and DG13S301, with alleles 4, 0, 2, 14 and 3, respectively (*p*-value of 1.02×10^{-7}). Carrier frequency of the haplotype is 7.3% in affected individuals and 0.3% in controls. These results are based on 437 patients and 721 controls. The area

that is common to all the haploytypes shown in the figure includes only one gene, FLAP.

FIG. 3.2 shows the alleles of the markers defining the most significant microsatellite marker haplotypes. The area defined with a black square is a common area to all the most significantly associated haplotypes. The FLAP nucleic acid is located between markers DG13S166 and D13S1238. Two marker haplotype involving alleles 0 and –2 for markers DG13S166 and S13S1238, respectively, is found in excess in patients. Carrier frequency of this haploype is 27% in patients and 15.4% in controls (*p*-value 1 X 10^{-3})

FIG. 4 shows the markers and genes around the FLAP (ALOX5AP) gene.

FIG. 5 shows the relative location of key SNPs and exons of the ALOX5AP/FLAP gene. Haplotype length varies between 33 to 68 kb.

FIGs. 6.1-6.82 shows the genomic sequence of the FLAP gene (SEQ ID NO: 1).

FIG. 7 shows the amino acid sequence of FLAP (SEQ ID NO: 2) and the mRNA of FLAP (SEQ ID NO: 3)

FIGs. 8.1-8.40 show the sequences of the FLAP nucleic acid flanking the SNPs that were identified by sequencing samples from patients (SEQ ID NOs: 398-535).

FIG. 9 shows a genome wide linkage scan using 1,000 microsatellite markers for all (black) (n=713), female (red), (n=140), male (blue) (n=575), and early onset MI patients (green) (n=194). The LOD score is expressed on the y axis and the distance from the pter in Kosambi cM on the x axis.

FIG. 10 shows a schematic view of the chromosome 13 linkage region showing the FLAP gene. (***a***) The linkage scan for female MI patients and the one LOD drop region that includes the FLAP gene; (***b***) Microsatellite association for all MI patients: single marker association (black dots) and two, three, four and five marker haplotype association (black, blue, green and red horizontal lines, respectively). The blue and the red arrows indicate the location of the most significant haplotype association across the FLAP gene in males and females, respectively. (*c*) The FLAP gene structure, with exons shown as colored cylinders, and the location of all the SNPs typed in the region (green vertical lines). The green

vertical lines indicate the position of the microsatellites (shown in *b*) and SNPs (shown in *c*) used in the analysis.

FIG. 11 shows linkage scan using framework microsatellite markers on chromosome 13 for male patients with ischemic stroke or TIA (n=342 in 164 families at 6 meiosis). The LOD score is expressed on the y axis and the distance from the pter in Kosambi cM on the x axis.

FIG. 12 shows pairwise linkage disequilibrium (LD) between SNPs in a 60 kb region encompassing FLAP. The markers are plotted equidistantly. Two measures of LD are shown: *D'* in the upper left triangle and *P* values in the lower right triangle. Colored lines indicate the positions of the exons of *FLAP* and the green stars indicate the location of the markers of the at-risk haplotype A4. Scales for the LD strength are provided for both measures to the right.

DETAILED DESCRIPTION OF THE INVENTION

Extensive genealogical information has been combined with powerful gene sharing methods to map a locus on chromosome 13q12 that is associated with myocardial infarction. Patients with myocardial infarction and controls were initially genotyped with microsatellite markers with an average spacing between markers of less than 100kb over the 12Mb candidate region. An epidemiological study of a population-based sample of MI patients demonstrated the relative risk for siblings of a female MI patient is significantly higher than the relative risk for siblings of a male proband (1.59 (CI 1.47 - 1.73) vs. 1.35 (CI 1.28 - 1.42)). The gender difference in risk of getting MI (males being more likely to get MI) also suggests somewhat different etiology between males and females, where MI in females might represent a more extreme phenotype. This study stratified the population according to sex to determine the genetic causes of MI for males and females. The results of the genome wide search of genes that cause MI in Iceland is described. This linkage analysis resulted in linkage on chromosome 13q12.

Initial haplotype association analysis studying 4 and 5 microsatellite marker haplotypes across the whole one lod drop region of the linkage peakshowed that the most significant haplotype in this region extended across the FLAP gene and was in

excess in patients, indicating that FLAP is a susceptibility gene for myocardial infarction. A region that was common to all the most significant microsatellite haplotypes included only one gene, the FLAP gene (see FIG. 3.1 and 3.2).

Subsequent studies revealed that a 4-SNP haplotype spanning the FLAP gene confers a near 2-fold risk of MI and stroke. In addition, male patients showed strongest association to the at-risk haplotype. Independent confirmation of FLAP association to MI was obtained in a British cohort of patients with sporadic MI. These findings indicate that FLAP is the first specific gene isolated that confers substantial risk of the complex traits of MI and stroke

The FLAP nucleic acid encodes a 5-lipoxygenase activating protein, which, in combination with 5-lipoxygenase (5-LO), is required for leukotriene synthesis. Inhibitors of its function impede translocation of 5-lipoxygenase from the cytoplasm to the cell membrane and inhibit activation of 5-lipoxygenase. One other member of the leukotriene pathway, CysLT2 receptor, maps to chromosome 13q14.2 (53 cM on FIG. 2). The region of this gene shows excess sharing identical by descent (LOD score=1) in female MI patients. This suggests that CysLT2 receptor might also play a role in the pathogenesis of MI. Mutations and/or polymorphisms within the FLAP nucleic acid show association with the disease and can be used for methods of diagnosis.

The product of the FLAP gene is involved in an important inflammatory pathway, and could thus be a predisposing factor for plaque rupture in MI. Since MI and stroke are both considered to be atherothrombotic diseases the MI at-risk haplotypes were assessed to determine whether they also were associated with stroke. Nineteen of the SNPs that defined the MI at-risk haplotypes (A and B series) were evaluated in stroke patients and unrelated controls. In the analysis, a subset of patients that did not have MI and were unrelated within 4 meiosis was used. The results from the haplotype association analysis are summarized in Table 8. The frequency of the at-risk haplotypes in all stroke patients was very similar to that of the MI patients and the haplotype conferred a similar relative risk. The B4 haplotype, previously described for MI, is carried by 19% of all stroke patients and 11% of controls. Carriers of this haplotype have nearly twofold risk (RR=1.95, P=1.6 10-4)

of having a stroke. Adding the fifth SNP (SG13S35) to the B4 haplotype increases the relative risk to 2.04 (p-value 5.8 10-5). The allelic frequency of this haplotype is 10.2% in stroke patients and 5.3% in controls. Also shown in Table 8 is a 4 SNP haplotype defined as Bs4 that is highly correlated with the B4 haplotype (r2=0.93). Bs4 haplotype has a RR of 2.01, carrier frequency in patients of 19% and population attributable risk of 10%. This haplotype was tested with different subtypes of stroke (Table 8). Of interest is that all stroke subtypes have a considerably higher frequency of the 'at-risk' haplotype than controls resulting in the increased relative risk.

NUCLEIC ACIDS OF THE INVENTION

FLAP Nucleic Acids, Portions and Variants

Accordingly, the invention pertains to isolated nucleic acid molecules comprising a human FLAP nucleic acid. The term, "FLAP nucleic acid," as used herein, refers to an isolated nucleic acid molecule encoding FLAP polypeptide. The FLAP nucleic acid molecules of the present invention can be RNA, for example, mRNA, or DNA, such as cDNA and genomic DNA. DNA molecules can be double-stranded or single-stranded; single stranded RNA or DNA can be either the coding, or sense strand or the non-coding, or antisense strand. The nucleic acid molecule can include all or a portion of the coding sequence of the gene or nucleic acid and can further comprise additional non-coding sequences such as introns and non-coding 3' and 5' sequences (including regulatory sequences, for example).

For example, a FLAP nucleic acid can consist of SEQ ID NOs: 1 or 3 or the complement thereof, or to a portion or fragment of such an isolated nucleic acid molecule (*e.g.*, cDNA or the nucleic acid) that encodes FLAP polypeptide (*e.g.*, a polypeptide such as SEQ ID NO: 2). In a preferred embodiment, the isolated nucleic acid molecule comprises a nucleic acid molecule selected from the group consisting of SEQ ID NOs: 1 or 3, or their complement thereof.

Additionally, the nucleic acid molecules of the invention can be fused to a marker sequence, for example, a sequence that encodes a polypeptide to assist in isolation or purification of the polypeptide. Such sequences include, but are not

limited to, those that encode a glutathione-S-transferase (GST) fusion protein and those that encode a hemagglutinin A (HA) polypeptide marker from influenza.

An “isolated” nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleic acid sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (*e.g.*, as in an RNA library). For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. In certain embodiments, an isolated nucleic acid molecule comprises at least about 50, 80 or 90% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term “isolated” also can refer to nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 5 kb, including but not limited to 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotides which flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of “isolated” as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. “Isolated” nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleic acid sequence can include a nucleic acid molecule or nucleic acid sequence that is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector is included in the definition of “isolated” as used herein.

Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. *In vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention are also encompassed by "isolated" nucleotide sequences. Such isolated nucleotide sequences are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (*e.g.*, from other mammalian species), for gene mapping (*e.g.*, by *in situ* hybridization with chromosomes), or for detecting expression of the nucleic acid in tissue (*e.g.*, human tissue), such as by Northern blot analysis.

The present invention also pertains to nucleic acid molecules which are not necessarily found in nature but which encode a FLAP polypeptide (*e.g.*, a polypeptide having an amino acid sequence comprising an amino acid sequence of SEQ ID NOs: 2), or another splicing variant of a FLAP polypeptide or polymorphic variant thereof. Thus, for example, DNA molecules that comprise a sequence that is different from the naturally occurring nucleic acid sequence but which, due to the degeneracy of the genetic code, encode a FLAP polypeptide of the present invention are also the subjects of this invention. The invention also encompasses nucleotide sequences encoding portions (fragments), or encoding variant polypeptides such as analogues or derivatives of a FLAP polypeptide. Such variants can be naturally occurring, such as in the case of allelic variation or single nucleotide polymorphisms, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides that can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the nucleotide (and/or resultant amino acid) changes are silent or conserved; that is, they do not alter the characteristics or activity of a FLAP polypeptide. In one preferred embodiment, the nucleotide sequences are fragments that comprise one or more polymorphic microsatellite markers. In another preferred embodiment, the nucleotide sequences are fragments that comprise one or more single nucleotide polymorphisms in a FLAP nucleic acid (*e.g.*, the single nucleotide polymorphisms set forth in Table 3, below).

Other alterations of the nucleic acid molecules of the invention can include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also pertains to nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleic acid sequence described herein (*e.g.*, nucleic acid molecules which specifically hybridize to a nucleic acid sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). In one embodiment, the invention includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleic acid sequence comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1 or 3 or the complement thereof. In another embodiment, the invention includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleic acid sequence encoding an amino acid sequence of SEQ ID NO: 2 or a polymorphic variant thereof. In a preferred embodiment, the variant that hybridizes under high stringency hybridizations has an activity of a FLAP.

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (*e.g.*, under high stringency conditions). "Specific hybridization," as used herein, refers to the ability of a first nucleic acid to hybridize to a second nucleic acid in a manner such that the first nucleic acid does not hybridize to any nucleic acid other than to the second nucleic acid (*e.g.*, when the first nucleic acid has a higher similarity to the second nucleic acid than to any other nucleic acid in a sample wherein the hybridization is to be performed). "Stringency conditions" for

hybridization is a term of art which refers to the incubation and wash conditions, *e.g.*, conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (*i.e.*, 100%) complementary to the second, or the first and second may share some degree of complementarity that is less than perfect (*e.g.*, 70%, 75%, 85%, 95%). For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions", "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6.3.6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), the entire teachings of which are incorporated by reference herein). The exact conditions which determine the stringency of hybridization depend not only on ionic strength (*e.g.*, 0.2X SSC, 0.1X SSC), temperature (*e.g.*, room temperature, 42°C, 68°C) and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules. Typically, conditions are used such that sequences at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95% or more identical to each other remain hybridized to one another. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (*e.g.*, selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology* 200: 546-556 (1991), and in, Ausubel, *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), which describes the determination of washing conditions for moderate or low stringency conditions.

Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of -17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2X SSC/0.1% SDS for 10 minutes at room temperature; a moderate stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2X SSC/0.1% SDS for 15 minutes at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1X SSC/0.1%SDS for 15 minutes at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

The percent homology or identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence for optimal alignment). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). When a position in one sequence is occupied by the same nucleotide or amino acid residue as the corresponding position in the other sequence, then the molecules are homologous at that position. As used herein, nucleic acid or amino acid "homology" is equivalent to nucleic acid or amino acid "identity". In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, for example, at least 40%, in certain embodiments at least 60%, and in other

embodiments at least 70%, 80%, 90% or 95% of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.*, *Nucleic Acids Res.* 25:389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, NBLAST) can be used. In one embodiment, parameters for sequence comparison can be set at score=100, wordlength=12, or can be varied (*e.g.*, W=5 or W=20).

Another preferred non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, *CABIOS* 4(1): 11-17 (1988). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package (Accelrys, Cambridge, UK). When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti, *Comput. Appl. Biosci.* 10:3-5 (1994); and FASTA described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444-8 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package using either a BLOSUM63 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. In yet another embodiment, the percent identity between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package using a gap weight of 50 and a length weight of 3.

The present invention also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleic acid sequence comprising SEQ ID NO: 1 or 3 or the complement of SEQ ID NO: 1 or 3, and also provides isolated nucleic acid molecules that contain a fragment

or portion that hybridizes under highly stringent conditions to a nucleic acid sequence encoding an amino acid sequence of the invention or polymorphic variant thereof. The nucleic acid fragments of the invention are at least about 15, for example, at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, encoding antigenic polypeptides described herein are particularly useful, such as for the generation of antibodies as described below.

Probes and Primers

In a related aspect, the nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid molecules. Such probes and primers include polypeptide nucleic acids, as described in Nielsen *et al.* (*Science* 254:1497-1500 (1991)).

A probe or primer comprises a region of nucleic acid that hybridizes to at least about 15, for example about 20-25, and in certain embodiments about 40, 50 or 75, consecutive nucleotides of a nucleic acid of the invention, such as a nucleic acid comprising a contiguous nucleic acid sequence of SEQ ID NOs: 1 or 3 or the complement of SEQ ID Nos: 1 or 3, or a nucleic acid sequence encoding an amino acid sequence of SEQ ID NO: 2 or polymorphic variant thereof. In preferred embodiments, a probe or primer comprises 100 or fewer nucleotides, in certain embodiments, from 6 to 50 nucleotides, for example, from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical to the contiguous nucleic acid sequence or to the complement of the contiguous nucleotide sequence, for example, at least 80% identical, in certain embodiments at least 90% identical, and in other embodiments at least 95% identical, or even capable of selectively hybridizing to the contiguous nucleic acid sequence or to the complement of the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, *e.g.*, radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

The nucleic acid molecules of the invention such as those described above can be identified and isolated using standard molecular biology techniques and the

sequence information provided herein. For example, nucleic acid molecules can be amplified and isolated using the polymerase chain reaction and synthetic oligonucleotide primers based on one or more of SEQ ID NOs: 1 or 3, or the complement thereof, or designed based on nucleotides based on sequences encoding one or more of the amino acid sequences provided herein. See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (Eds. Innis *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucl. Acids Res.* 19:4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1:17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202. The nucleic acid molecules can be amplified using cDNA, mRNA or genomic DNA as a template, cloned into an appropriate vector and characterized by DNA sequence analysis.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989), Landegren *et al.*, *Science* 241:1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA* 87:1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be labeled, for example, radiolabeled, and used as a probe for screening a cDNA library derived from human cells, mRNA in zap express, ZIPLOX or other suitable vector. Corresponding clones can be isolated, DNA can obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. For example, the direct analysis of the nucleic acid molecules of the present invention can be accomplished using well-known methods that are commercially available. See, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New

York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)). Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced and further characterized.

Antisense nucleic acid molecules of the invention can be designed using the nucleotide sequences of SEQ ID NOs: 1 or 3 and/or the complement of one or more of SEQ ID NOs: 1 or 3 and/or a portion of one or more of SEQ ID NOs: 1 or 3 or the complement of one or more of SEQ ID NOs: 1 or 3 and/or a sequence encoding the amino acid sequences of SEQ ID NOs: 2 or encoding a portion of one or more of SEQ ID NOs: 1 or 3 or their complement. They can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid molecule can be produced biologically using an expression vector into which a nucleic acid molecule has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid molecule will be of an antisense orientation to a target nucleic acid of interest).

The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify one or more of the disorders related to FLAP, and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample. The nucleic acid sequences can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using DNA immunization techniques, and as an antigen to raise anti-DNA antibodies or elicit immune responses. Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions or nucleic acid regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a

biological sample. Additionally, the nucleotide sequences of the invention can be used to identify and express recombinant polypeptides for analysis, characterization or therapeutic use, or as markers for tissues in which the corresponding polypeptide is expressed, either constitutively, during tissue differentiation, or in diseased states. The nucleic acid sequences can additionally be used as reagents in the screening and/or diagnostic assays described herein, and can also be included as components of kits (*e.g.*, reagent kits) for use in the screening and/or diagnostic assays described herein.

Vectors

Another aspect of the invention pertains to nucleic acid constructs containing a nucleic acid molecule of SEQ ID NOs: 1 or 3 or the complement thereof (or a portion thereof). Yet another aspect of the invention pertains to nucleic acid constructs containing a nucleic acid molecule encoding an amino acid of SEQ ID NO: 2 or polymorphic variant thereof. The constructs comprise a vector (*e.g.*, an expression vector) into which a sequence of the invention has been inserted in a sense or antisense orientation. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, such as expression vectors, are capable of directing the expression of genes or nucleic acids to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. However, the invention is intended to include such other forms of expression vectors, such as

viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" or "operatively linked" is intended to mean that the nucleic acid sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleic acid sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, "Gene Expression Technology", *Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleic acid sequence in many types of host cell and those which direct expression of the nucleic acid sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed and the level of expression of polypeptide desired. The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides, including fusion polypeptides, encoded by nucleic acid molecules as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid molecule of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing a foreign nucleic acid molecule (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene or nucleic acid that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene or nucleic acid of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector as the nucleic acid molecule of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be

identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene or nucleic acid will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic host cell or eukaryotic host cell in culture can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a nucleic acid molecule of the invention has been introduced (*e.g.*, an exogenous FLAP nucleic acid, or an exogenous nucleic acid encoding a FLAP polypeptide). Such host cells can then be used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into the genome or homologous recombinant animals in which endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleic acid sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal include a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens and amphibians. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA

molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Pat. No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, *Current Opinion in BioTechnology* 2:823-829 (1991) and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169. Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.*, *Nature* 385:810-813 (1997) and PCT Publication Nos. WO 97/07668 and WO 97/07669.

POLYPEPTIDES OF THE INVENTION

The present invention also pertains to isolated polypeptides encoded by FLAP nucleic acids ("FLAP polypeptides"), and fragments and variants thereof, as well as polypeptides encoded by nucleotide sequences described herein (*e.g.*, other splicing variants). The term "polypeptide" refers to a polymer of amino acids, and not to a specific length; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell (*e.g.*, in a "fusion protein") and still be "isolated" or "purified."

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity. In one

embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When a polypeptide is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the polypeptide preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, a polypeptide of the invention comprises an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 or 3, or the complement of SEQ ID NO: 1 or 3, or portions thereof, or a portion or polymorphic variant thereof. However, the polypeptides of the invention also encompass fragment and sequence variants. Variants include a substantially homologous polypeptide encoded by the same genetic locus in an organism, *i.e.*, an allelic variant, as well as other splicing variants. Variants also encompass polypeptides derived from other genetic loci in an organism, but having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1 or 3 or their complement, or portions thereof, or having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of nucleotide sequences encoding SEQ ID NO: 2 or polymorphic variants thereof. Variants also include polypeptides substantially homologous or identical to these polypeptides but derived from another organism, *i.e.*, an ortholog. Variants also include polypeptides that are substantially

homologous or identical to these polypeptides that are produced by chemical synthesis. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by recombinant methods.

As used herein, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 45-55%, in certain embodiments at least about 70-75%, and in other embodiments at least about 80-85%, and in others greater than about 90% or more homologous or identical. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule hybridizing to SEQ ID NO: 1 or 3 or portion thereof, under stringent conditions as more particularly described above, or will be encoded by a nucleic acid molecule hybridizing to a nucleic acid sequence encoding SEQ ID NO: 2 or a portion thereof or polymorphic variant thereof, under stringent conditions as more particularly described thereof.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by a polypeptide encoded by a nucleic acid molecule of the invention. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Further, variant polypeptides can be fully functional or can lack function in one or more activities. Fully functional variants typically contain

only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al., Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity *in vitro*, or *in vitro* proliferative activity. Sites that are critical for polypeptide activity can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al., Science 255*:306-312 (1992)).

The invention also includes fragments of the polypeptides of the invention. Fragments can be derived from a polypeptide encoded by a nucleic acid molecule comprising SEQ ID NO: 1 or 3, or the complement of SEQ ID NO: 1 or 3 (or other variants). However, the invention also encompasses fragments of the variants of the polypeptides described herein. As used herein, a fragment comprises at least 6 contiguous amino acids. Useful fragments include those that retain one or more of the biological activities of the polypeptide as well as fragments that can be used as an immunogen to generate polypeptide-specific antibodies.

Biologically active fragments (peptides which are, for example, 6, 9, 12, 15, 16, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can comprise a domain, segment, or motif that has been identified by analysis of the polypeptide sequence using well-known methods, *e.g.*, signal peptides, extracellular domains, one or more transmembrane segments or loops, ligand binding regions, zinc finger domains, DNA binding domains, acylation sites, glycosylation sites, or phosphorylation sites.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion polypeptides. These comprise a polypeptide of the invention operatively linked to a heterologous protein or polypeptide having an amino acid sequence not substantially homologous to the polypeptide. "Operatively linked" indicates that the polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the polypeptide. In one embodiment the fusion polypeptide does not affect function of the polypeptide *per se*. For example, the fusion polypeptide can be a GST-fusion polypeptide in which the polypeptide sequences are fused to the C-terminus of the GST sequences. Other types of fusion polypeptides include, but are not limited to, enzymatic fusion polypeptides, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion polypeptides, particularly poly-His fusions, can facilitate the purification of recombinant polypeptide. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a polypeptide can be increased using a heterologous signal sequence. Therefore, in another embodiment, the fusion polypeptide contains a heterologous signal sequence at its N-terminus.

EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.*, *Journal of Molecular Recognition, 8*:52-58 (1995) and Johanson *et al.*, *The Journal of Biological Chemistry, 270*,16:9459-9471 (1995). Thus, this invention also encompasses soluble fusion polypeptides containing a polypeptide of the invention

and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE).

A chimeric or fusion polypeptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of nucleic acid fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive nucleic acid fragments which can subsequently be annealed and re-amplified to generate a chimeric nucleic acid sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST protein). A nucleic acid molecule encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide.

The isolated polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. In one embodiment, the polypeptide is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the polypeptide expressed in the host cell. The polypeptide can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

The polypeptides of the present invention can be used to raise antibodies or to elicit an immune response. The polypeptides can also be used as a reagent, *e.g.*, a labeled reagent, in assays to quantitatively determine levels of the polypeptide or a molecule to which it binds (*e.g.*, a ligand) in biological fluids. The polypeptides can also be used as markers for cells or tissues in which the corresponding polypeptide is preferentially expressed, either constitutively, during tissue differentiation, or in diseased states. The polypeptides can be used to isolate a corresponding binding agent, *e.g.*, ligand, such as, for example, in an interaction trap assay, and to screen for

peptide or small molecule antagonists or agonists of the binding interaction. For example, because members of the leukotriene pathway including FLAP bind to receptors, the leukotriene pathway polypeptides can be used to isolate such receptors.

ANTIBODIES OF THE INVENTION

Polyclonal and/or monoclonal antibodies that specifically bind one form of the polypeptide or nucleic acid product (*e.g.*, a polypeptide encoded by a nucleic acid having a SNP as set forth in Table 3), but not to another form of the polypeptide or nucleic acid product, are also provided. Antibodies are also provided which bind a portion of either polypeptide encoded by nucleic acids of the invention (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3, or the complement of SEQ ID NO: 1 or SEQ ID NO: 3), or to a polypeptide encoded by nucleic acids of the invention that contain a polymorphic site or sites. The invention also provides antibodies to the polypeptides and polypeptide fragments of the invention, or a portion thereof, or having an amino acid sequence encoded by a nucleic acid molecule comprising all or a portion of SEQ ID NOs: 1 or 3, or the complement thereof, or another variant or portion thereof. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds an antigen. A molecule that specifically binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind to a polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, polypeptide of the invention or fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature* 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al.*, *Immunol. Today* 4:72 (1983)); the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, Inc., pp. 77-96); or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan *et al.* (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, *Current Protocols in Immunology, supra*; Galfre *et al.*, *Nature* 266:55052 (1977); R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.* 54:387-402 (1981). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to a polypeptide of the invention can be identified and isolated

by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *Surf*ZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, *Bio/Technology* 9: 1370-1372 (1991); Hay *et al.*, *Hum. Antibod. Hybridomas* 3:81-85 (1992); Huse *et al.*, *Science* 246:1275-1281 (1989); Griffiths *et al.*, *EMBO J.* 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate a polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for a polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials,

luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, ß-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

DIAGNOSTIC ASSAYS

DIAGNOSIS USING PROBES, PRIMERS, POLYPEPTIDES, AND ANTIBODIES

The nucleic acids, probes, primers, polypeptides and antibodies described herein can be used in methods of diagnosis of a susceptibility to MI or stroke, or to a disease or condition associated with a gene such as FLAP, as well as in kits useful for diagnosis of a susceptibility to MI or stroke, or to a disease or condition associated with FLAP. In one embodiment, the kit useful for diagnosis of susceptibility to MI or stroke, or to a disease or condition associated with FLAP comprises primers as described herein, wherein the primers contain one or more of the SNPs identified in Table 3.

In one embodiment of the invention, diagnosis of susceptibility to MI or stroke (or diagnosis of or susceptibility to a disease or condition associated with FLAP), is made by detecting a polymorphism in a FLAP nucleic acid as described herein. The polymorphism can be an alteration in a FLAP nucleic acid, such as the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift alteration; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an

interruption of the coding sequence of the gene or nucleic acid; duplication of all or a part of the gene or nucleic acid; transposition of all or a part of the gene or nucleic acid; or rearrangement of all or a part of the gene or nucleic acid. More than one such alteration may be present in a single gene or nucleic acid. Such sequence changes cause an alteration in the polypeptide encoded by a FLAP nucleic acid. For example, if the alteration is a frame shift alteration, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a disease or condition associated with a FLAP nucleic acid or a susceptibility to a disease or condition associated with a FLAP nucleic acid can be a synonymous alteration in one or more nucleotides (*i.e.*, an alteration that does not result in a change in the polypeptide encoded by a FLAP nucleic acid). Such a polymorphism may alter splicing sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the nucleic acid. A FLAP nucleic acid that has any of the alteration described above is referred to herein as an "altered nucleic acid."

In a first method of diagnosing a susceptibility to MI or stroke, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used (see *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 1999). For example, a biological sample from a test subject (a "test sample") of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having, being susceptible to or predisposed for, or carrying a defect for, a susceptibility to a disease or condition associated with a FLAP nucleic acid (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source which contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in a nucleic acid is present, and/or to determine which splicing variant(s) encoded by

the FLAP is present. The presence of the polymorphism or splicing variant(s) can be indicated by hybridization of the nucleic acid in the genomic DNA, RNA, or cDNA to a nucleic acid probe. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe can contain at least one polymorphism in a FLAP nucleic acid or contains a nucleic acid encoding a particular splicing variant of a FLAP nucleic acid. The probe can be any of the nucleic acid molecules described above (*e.g.*, the nucleic acid, a fragment, a vector comprising the nucleic acid, a probe or primer, etc.).

To diagnose a susceptibility to MI or stroke (or a disease or condition associated with FLAP), the test sample containing a FLAP nucleic acid is contacted with at least one nucleic acid probe to form a hybridization sample. A preferred probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be all or a portion of one of SEQ ID NOs: 1 and 3, or the complement thereof or a portion thereof; or can be a nucleic acid encoding all or a portion of one of SEQ ID NO: 2. Other suitable probes for use in the diagnostic assays of the invention are described above (see *e.g.*, probes and primers discussed under the heading, "Nucleic Acids of the Invention").

The hybridization sample is maintained under conditions that are sufficient to allow specific hybridization of the nucleic acid probe to a FLAP nucleic acid. "Specific hybridization", as used herein, indicates exact hybridization (*e.g.*, with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and FLAP nucleic acid in the test sample, then the FLAP has the polymorphism, or is the splicing variant,

that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a polymorphism in the FLAP nucleic acid, or of the presence of a particular splicing variant encoding the FLAP nucleic acid, and is therefore diagnostic for a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke).

In Northern analysis (see *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*) the hybridization methods described above are used to identify the presence of a polymorphism or a particular splicing variant, associated with a disease or condition associated with or a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke). For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a polymorphism in a FLAP nucleic acid, or of the presence of a particular splicing variant encoded by a FLAP nucleic acid, and is therefore diagnostic for the disease or condition associated with FLAP, or for susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke).

For representative examples of use of nucleic acid probes, see, for example, U.S. Patents No. 5,288,611 and 4,851,330. Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P.E. *et al.*, *Bioconjugate Chemistry* 5, American Chemical Society, p. 1 (1994). The PNA probe can be designed to specifically hybridize to a nucleic acid having a polymorphism associated with a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke). Hybridization of the PNA probe to a FLAP nucleic acid as described herein is diagnostic for the disease or condition or the susceptibility to the disease or condition.

In another method of the invention, mutation analysis by restriction digestion can be used to detect an altered nucleic acid, or nucleic acids containing a

polymorphism(s), if the mutation or polymorphism in the nucleic acid results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify a FLAP nucleic acid (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described (see *Current Protocols in Molecular Biology*, *supra*). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the alteration or polymorphism in the FLAP nucleic acid, and therefore indicates the presence or absence of the susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke).

Sequence analysis can also be used to detect specific polymorphisms in the FLAP nucleic acid. A test sample of DNA or RNA is obtained from the test individual. PCR or other appropriate methods can be used to amplify the nucleic acid, and/or its flanking sequences, if desired. The sequence of a FLAP nucleic acid, or a fragment of the nucleic acid, or cDNA, or fragment of the cDNA, or mRNA, or fragment of the mRNA, is determined, using standard methods. The sequence of the nucleic acid, nucleic acid fragment, cDNA, cDNA fragment, mRNA, or mRNA fragment is compared with the known nucleic acid sequence of the nucleic acid, cDNA (*e.g.*, one or more of SEQ ID NOs: 1 or 3, and/or the complement of SEQ ID NO: 1 or 3), or a nucleic acid sequence encoding SEQ ID NO: 2 or a fragment thereof) or mRNA, as appropriate. The presence of a polymorphism in the FLAP nucleic acid indicates that the individual has disease or a susceptibility to a disease associated with FLAP (*e.g.*, MI or stroke).

Allele-specific oligonucleotides can also be used to detect the presence of polymorphism(s) in the FLAP nucleic acid, through the use of dot-blot hybridization of amplified oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, *Nature* 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, for example, approximately 15-30 base pairs, that specifically hybridizes to a FLAP nucleic acid, and that contains a polymorphism associated with a susceptibility to a disease or condition associated

with FLAP (*e.g.*, MI or stroke). An allele-specific oligonucleotide probe that is specific for particular polymorphisms in a FLAP nucleic acid can be prepared, using standard methods (see *Current Protocols in Molecular Biology*, *supra*). To identify polymorphisms in the nucleic acid associated with disease or susceptibility to disease, a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of a FLAP nucleic acid, and its flanking sequences. The DNA containing the amplified FLAP nucleic acid (or fragment of the nucleic acid) is dot-blotted, using standard methods (see *Current Protocols in Molecular Biology*, *supra*), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified FLAP is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a polymorphism in the FLAP, and is therefore indicative of a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke).

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, *e.g.*, WO 93/22456).

With the addition of such analogs as locked nucleic acids (LNAs), the size of primers and probes can be reduced to as few as 8 bases. LNAs are a novel class of bicyclic DNA analogs in which the 2' and 4' positions in the furanose ring are joined via an O-methylene (oxy-LNA), S-methylene (thio-LNA), or amino methylene (amino-LNA) moiety. Common to all of these LNA variants is an affinity toward complementary nucleic acids, which is by far the highest reported for a DNA analog.

For example, particular all oxy-LNA nonamers have been shown to have melting temperatures of 64°C and 74°C when in complex with complementary DNA or RNA, respectively, as oposed to 28°C for both DNA and RNA for the corresponding DNA nonamer. Substantial increases in T_m are also obtained when LNA monomers are used in combination with standard DNA or RNA monomers. For primers and probes, depending on where the LNA monomers are included (*e.g.*, the 3' end, the 5'end, or in the middle), the T_m could be increased considerably

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual, can be used to identify polymorphisms in a FLAP nucleic acid. For example, in one embodiment, an oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "Genechips™," have been generally described in the art, for example, U.S. Pat. No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and WO 92/10092. These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor *et al.*, *Science* 251:767-777 (1991); Pirrung *et al.*, U.S. Pat. 5,143,854; (see also PCT Application WO 90/15070); Fodor *et al.*, PCT Publication WO 92/10092; and U.S. Pat. 5,424,186, the entire teachings of each of which are incorporated by reference herein. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Pat. 5,384,261, the entire teachings of which are incorporated by reference herein. In another example, linear arrays can be utilized.

Once an oligonucleotide array is prepared, a nucleic acid of interest is hybridized with the array and scanned for polymorphisms. Hybridization and scanning are generally carried out by methods described herein and also in, *e.g.*, published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Pat. No. 5,424,186, the entire teachings of which are incorporated by reference herein. In brief, a target nucleic acid sequence that includes one or more previously identified

polymorphic markers is amplified using well-known amplification techniques, *e.g.*, PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate conditions. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array. In a reverse method, a probe, containing a polymorphism, can be coupled to a solid surface and PCR amplicons are then added to hybridize to these probes.

Although primarily described in terms of a single detection block, *e.g.*, detection of a single polymorphism arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. It will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

Additional uses of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patents Nos. 5,858,659 and 5,837,832, the entire teachings of which are incorporated by reference herein. Other methods of nucleic acid analysis can be used to detect polymorphisms in a nucleic acid described herein, or variants encoded by a nucleic acid described herein. Representative methods include direct manual sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (1988); Sanger, F. *et al.*, *Proc. Natl. Acad. Sci., USA* 74:5463-5467 (1977); Beavis *et al.* U.S. Pat. No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V.C. *et al.*, *Proc. Natl. Acad. Sci. USA* 86:232-236 (1989)), mobility shift analysis

(Orita, M. *et al., Proc. Natl. Acad. Sci. USA* 86:2766-2770 (1989)), restriction enzyme analysis (Flavell *et al., Cell* 15:25 (1978); Geever, *et al., Proc. Natl. Acad. Sci. USA* 78:5081 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton *et al., Proc. Natl. Acad. Sci. USA* 85:4397-4401 (1985)); RNase protection assays (Myers, R.M. *et al., Science* 230:1242 (1985)); use of polypeptides which recognize nucleotide mismatches, such as *E. coli* mutS protein; allele-specific PCR, for example.

In one embodiment of the invention, diagnosis of a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke) can also be made by expression analysis by quantitative PCR (kinetic thermal cycling). This technique utilizing TaqMan® can be used to allow the identification of polymorphisms and whether a patient is homozygous or heterozygous. The technique can assess the presence of an alteration in the expression or composition of the polypeptide encoded by a FLAP nucleic acid or splicing variants encoded by a FLAP nucleic acid. Further, the expression of the variants can be quantified as physically or functionally different.

In another embodiment of the invention, diagnosis of a susceptibility to MI or stroke (or of another disease or condition associated with FLAP) can also be made by examining expression and/or composition of a FLAP polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. A test sample from an individual is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by a FLAP nucleic acid, or for the presence of a particular variant encoded by a FLAP nucleic acid. An alteration in expression of a polypeptide encoded by a FLAP nucleic acid can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by a FLAP nucleic acid is an alteration in the qualitative polypeptide expression (*e.g.*, expression of an altered FLAP polypeptide or of a different splicing variant). In a preferred embodiment, diagnosis of disease or condition associated with FLAP or a susceptibility to a disease or condition associated with FLAP is made by detecting a particular splicing variant encoded by that FLAP variant, or a particular pattern of splicing variants.

Both such alterations (quantitative and qualitative) can also be present. An "alteration" in the polypeptide expression or composition, refers to an alteration in expression or composition in a test sample, as compared with the expression or composition of polypeptide by a FLAP nucleic acid in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is from the same type of cells), and is from an individual who is not affected by the disease or a susceptibility to a disease or condition associated with a FLAP nucleic acid. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, is indicative of disease or condition associated with FLAP or a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke). Similarly, the presence of one or more different splicing variants in the test sample, or the presence of significantly different amounts of different splicing variants in the test sample, as compared with the control sample, is indicative of a susceptibility to a disease or condition associated with a FLAP nucleic acid. Various means of examining expression or composition of the polypeptide encoded by a FLAP nucleic acid can be used, including: spectroscopy, colorimetry, electrophoresis, isoelectric focusing and immunoassays (*e.g.*, David *et al.*, U.S. Pat. 4,376,110) such as immunoblotting (see also *Current Protocols in Molecular Biology*, particularly Chapter 10). For example, in one embodiment, an antibody capable of binding to the polypeptide (*e.g.*, as described above), preferably an antibody with a detectable label, can be used. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. Western blotting analysis, using an antibody as described above that specifically binds to a polypeptide encoded by an altered FLAP (*e.g.*, by a FLAP having a SNP as shown in Table 3), or an antibody

that specifically binds to a polypeptide encoded by a non-altered nucleic acid, or an antibody that specifically binds to a particular splicing variant encoded by a nucleic acid, can be used to identify the presence in a test sample of a particular splicing variant or of a polypeptide encoded by a polymorphic or altered FLAP, or the absence in a test sample of a particular splicing variant or of a polypeptide encoded by a non-polymorphic or non-altered nucleic acid. The presence of a polypeptide encoded by a polymorphic or altered nucleic acid, or the absence of a polypeptide encoded by a non-polymorphic or non-altered nucleic acid, is diagnostic for a susceptibility to a disease or condition associated with FLAP, as is the presence (or absence) of particular splicing variants encoded by the FLAP nucleic acid.

In one embodiment of this method, the level or amount of polypeptide encoded by a FLAP nucleic acid in a test sample is compared with the level or amount of the polypeptide encoded by the FLAP in a control sample. A level or amount of the polypeptide in the test sample that is higher or lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the polypeptide encoded by the FLAP, and is diagnostic for disease or condition, or for a susceptibility to a disease or condition, associated with that FLAP. Alternatively, the composition of the polypeptide encoded by a FLAP nucleic acid in a test sample is compared with the composition of the polypeptide encoded by the FLAP in a control sample (*e.g.*, the presence of different splicing variants). A difference in the composition of the polypeptide in the test sample, as compared with the composition of the polypeptide in the control sample, is diagnostic for a susceptibility to a disease or condition associated with that FLAP. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample. A difference in the amount or level of the polypeptide in the test sample, compared to the control sample; a difference in composition in the test sample, compared to the control sample; or both a difference in the amount or level, and a difference in the composition, is indicative of a susceptibility to a disease or condition, associated with FLAP (*e.g.*, MI or stroke).

DIAGNOSIS UTILIZING AT-RISK HAPLOTYPES

The invention further pertains to a method for the diagnosis and identification of susceptibility to myocardial infarction or stroke in an individual, by identifying an at-risk haplotype in FLAP. As used herein, combinations of genetic markers are referred to herein as "haplotypes," and the present invention describes methods whereby detection of particular haplotypes is indicative of a susceptibility to myocardial infarction or stroke. The detection of the particular genetic markers that make up the particular haplotypes can be performed by a variety of methods described herein and known in the art. For example, genetic markers can be detected at the nucleic acid level, *e.g.*, by direct sequencing or at the amino acid level if the genetic marker affects the coding sequence of FLAP, *e.g.*, by immunoassays based on antibodies that recognize the FLAP protein or a particular FLAP variant protein.

In one embodiment of the invention, diagnosis of a susceptibility to MI or stroke is made by detecting a haplotype associated with FLAP as described herein. The FLAP-associated haplotypes (*e.g.*, those described in Tables 4, 5 and 13), describe a set of genetic markers ("alleles") associated with FLAP. In a certain embodiment, the haplotype can comprise one or more alleles, two or more alleles, three or more alleles, four or more alleles, or five or more alleles. The genetic markers are particular "alleles" at "polymorphic sites" associated with FLAP. A nucleotide position at which more than one sequence is possible in a population (either a natural population or a synthetic population, *e.g.*, a library of synthetic molecules), is referred to herein as a "polymorphic site". Where a polymorphic site is a single nucleotide in length, the site is referred to as a single nucleotide polymorphism ("SNP"). For example, if at a particular chromosomal location, one member of a population has an adenine and another member of the population has a thymine at the same position, then this position is a polymorphic site, and, more specifically, the polymorphic site is a SNP. Polymorphic sites can allow for differences in sequences based on substitutions, insertions or deletions. Each version of the sequence with respect to the polymorphic site is referred to herein as an "allele" of the polymorphic site. Thus, in the previous example, the SNP allows for both an adenine allele and a thymine allele.

Typically, a reference sequence is referred to for a particular sequence. Alleles that differ from the reference are referred to as "variant" alleles. For example, the reference FLAP sequence is described herein by SEQ ID NO: 1(genomic) or SEQ ID NO: 3 (mRNA). The term, "variant FLAP", as used herein, refers to a FLAP sequence that differs from SEQ ID NO: 1 or SEQ ID NO: 3, but is otherwise substantially similar. The genetic markers that make up the haplotypes described herein include FLAP variants. The variants of FLAP that are used to determine the haplotypes disclosed herein of the present invention are associated with a susceptibility to MI or stroke.Additional variants can include changes that affect a FLAP polypeptide, as described above.

Haplotypes are a combination of genetic markers, *e.g.*, particular alleles at polymorphic sites. The haplotypes described herein (*e.g.*, in Table 4 or 5; haplotypes B4, B5, B6, A4, A5; HapB) are found more frequently in individuals having MI and/or stroke than in individuals not affected by these diseases. Therefore, these haplotypes have predictive value for detecting susceptibility to MI or stroke in an individual.

In one embodiment, the at-risk haplotype is one which confers a significant risk of MI or stroke. In one embodiment, significance associated with a haplotype is measured by an odds ratio. In a further embodiment, the significance is measured by a percentage. In one embodiment, a significant risk is measured as an odds ratio of at least about 1.2, including by not limited to: 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 1.9. In a further embodiment, an odds ratio of at least 1.2 is significant. In a further embodiment, an odds ratio of at least about 1.5 is significant. In a further embodiment, a significant increase in risk is at least about 1.7 is significant. In a further embodiment, a significant increase in risk is at least about 20%, including but not limited to about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, and 98%. In a further embodiment, a significant increase in risk is at least about 50%. It is understood however, that identifying whether a risk is medically significant may also depend on a variety of factors, including the specific disease, the haplotype, and often, environmental factors.

The invention also pertains to methods of diagnosing a susceptibility to myocardial infarction or stroke in an individual, comprising screening for an at-risk haplotype in the FLAP nucleic acid that is more frequently present in an individual susceptible to myocardial infarction or stroke (affected), compared to the frequency of its presence in a healthy individual (control), wherein the presence of the haplotype is indicative of susceptibility to myocardial infarction or stroke. As an example of a simple test for correlation would be a Fisher-exact test on a two by two table. Given a cohort of chromosomes the two by two table is constructed out of the number of chromosomes that include both of the haplotypes, one of the haplotype but not the other and neither of the haplotypes.

In certain embodiments, the screening for the presence of an at-risk haplotype comprises screening for an at-risk haplotype within or near FLAP that significantly correlates with a haplotype such as a halotype shown in Table 4; a haplotype shown in Table 5; a haplotype shown in Table 13; haplotype B4; haplotype B5; haplotype B6; haplotype A4; haplotype A5; or haplotype HapB. In other embodiments, screening for the presence of an at-risk haplotype comprises screening for an at-risk haplotype within or near FLAP that significantly correlates with susceptibility to myocardial infarction or stroke.

In one particular embodiment, the at-risk haplotype is characterized by the presence of polymorphism(s) represented in Table 3. For example, DG00AAFIU, where the SNP can be a "C" or a "T"; SG13S25, where the SNP can be a "G" or an "A"; DG00AAJFF, where the SNP can be a "G" or an "A"; DG00AAHII, where the SNP can be a "G" or an "A"; DG00AAHID, where the SNP can be a "T" or an "A"; B_SNP_310657, where the SNP can be a "G" or an "A"; SG13S30, where the SNP can be a "G" or a "T"; SG13S32, where the SNP can be a "C" or an "A"; SG13S42, where the SNP can be a "G" or an "A"; and SG13S35, where the SNP can be a "G" or an "A". In another embodiment, the at-risk haplotype is selected from the group consisting of: haplotype B4, B5, B6, A4 and A5. The at-risk haplotype can also comprise a combination of the markers in the haplotypes B4, B5, B6, A4 and/or A5. In further embodiments, the at-risk haplotype can be haplotype HapB. In other

embodiments, the at-risk haplotype comprises a polymorphism shown in Table 3 and/or in Table 13.

Standard techniques for genotyping for the presence of SNPs and/or microsatellite markers that are associated with myocardial infarctionor stroke can be used, such as fluorescent based techniques (Chen, *et al.*, *Genome Res. 9*, 492 (1999), PCR, LCR, Nested PCR and other techniques for nucleic acid amplification. In a preferred embodiment, the method comprises assessing in an individual the presence or frequency of SNPs and/or microsatellites in the FLAP nucleic acid that are associated with myocardial infarction or stroke, wherein an excess or higher frequency of the SNPs and/or microsatellites compared to a healthy control individual is indicative that the individual is susceptible to myocardial infarction or stroke.

Haplotype analysis involves defining a candidate susceptibility locus using LOD scores. The defined regions are then ultra-fine mapped with microsatellite markers with an average spacing between markers of less than 100kb. All usable microsatellite markers that found in public databases and mapped within that region can be used. In addition, microsatellite markers identified within the deCODE genetics sequence assembly of the human genome can be used.

The frequencies of haplotypes in the patient and the control groups using an expectation-maximization algorithm can be estimated (Dempster A. *et al.*, 1977. *J. R. Stat. Soc. B*, 39:1-389). An implementation of this algorithm that can handle missing genotypes and uncertainty with the phase can be used. Under the null hypothesis, the patients and the controls are assumed to have identical frequencies. Using a likelihood approach, an alternative hypothesis where a candidate at-risk-haplotype, which can include the FLAP SNPs, is allowed to have a higher frequency in patients than controls, while the ratios of the frequencies of other haplotypes are assumed to be the same in both groups is tested. Likelihoods are maximized separately under both hypotheses and a corresponding 1-df likelihood ratio statistic is used to evaluate the statistic significance.

To look for at-risk-haplotypes in the 1-lod drop, for example, association of all possible combinations of genotyped markers is studied, provided those markers span a practical region. The combined patient and control groups can be randomly divided

into two sets, equal in size to the original group of patients and controls. The haplotype analysis is then repeated and the most significant p-value registered is determined. This randomization scheme can be repeated, for example, over 100 times to construct an empirical distribution of p-values. In a preferred embodiment, a p-value of <0.05 is indicative of an at-risk haplotype.

A detailed discussion of haplotype analysis follows.

Haplotype analysis

Our general approach to haplotype analysis involves using likelihood-based inference applied to NEsted MOdels. The method is implemented in our program NEMO, which allows for many polymorphic markers, SNPs and microsatellites. The method and software are specifically designed for case-control studies where the purpose is to identify haplotype groups that confer different risks. It is also a tool for studying LD structures.

When investigating haplotypes constructed from many markers, apart from looking at each haplotype individually, meaningful summaries often require putting haplotypes into groups. A particular partition of the haplotype space is a model that assumes haplotypes within a group have the same risk, while haplotypes in different groups can have different risks. Two models/partitions are nested when one, the alternative model, is a finer partition compared to the other, the null model, *i.e,* the alternative model allows some haplotypes assumed to have the same risk in the null model to have different risks. The models are nested in the classical sense that the null model is a special case of the alternative model. Hence traditional generalized likelihood ratio tests can be used to test the null model against the alternative model. Note that, with a multiplicative model, if haplotypes h_i and h_j are assumed to have the same risk, it corresponds to assuming that $f_i/p_i = f_j/p_j$ where f and p denote haplotype frequencies in the affected population and the control population respectively.

One common way to handle uncertainty in phase and missing genotypes is a two-step method of first estimating haplotype counts and then treating the estimated counts as the exact counts, a method that can sometimes be problematic (*e.g.,* see the information measure section below) and may require randomization to properly

evaluate statistical significance. In NEMO, maximum likelihood estimates, likelihood ratios and p-values are calculated directly, with the aid of the EM algorithm, for the observed data treating it as a missing-data problem.

NEMO allows complete flexibility for partitions. For example, the first haplotype problem described in the Methods section on Statistical analysis considers testing whether h_1 has the same risk as the other haplotypes $h_2, \ldots, h_k$. Here the alternative grouping is $[h_1], [h_2, \ldots, h_k]$ and the null grouping is $[h_1, \ldots, h_k]$. The second haplotype problem in the same section involves three haplotypes h_1 = G0, h_2 = GX and h_3 = AX, and the focus is on comparing h_1 and h_2. The alternative grouping is $[h_1], [h_2], [h_3]$ and the null grouping is $[h_1, h_2], [h_3]$. If composite alleles exist, one could collapse these alleles into one at the data processing stage, and performed the test as described. This is a perfectly valid approach, and indeed, whether we collapse or not makes no difference if there were no missing information regarding phase. But, with the actual data, if each of the alleles making up a composite correlates differently with the SNP alleles, this will provide some partial information on phase. Collapsing at the data processing stage will unnecessarily increase the amount of missing information. A nested-models/partition framework can be used in this scenario. Let h_2 be split into $h_{2a}, h_{2b}, \ldots, h_{2e}$, and h_3 be split into $h_{3a}, h_{3b}, \ldots, h_{3e}$. Then the alternative grouping is $[h_1], [h_{2a}, h_{2b}, \ldots, h_{2e}], [h_{3a}, h_{3b}, \ldots, h_{3e}]$ and the null grouping is $[h_1, h_{2a}, h_{2b}, \ldots, h_{2e}], [h_{3a}, h_{3b}, \ldots, h_{3e}]$. The same method can be used to handle composite where collapsing at the data processing stage is not even an option since L_C represents multiple haplotypes constructed from multiple SNPs. Alternatively, a 3-way test with the alternative grouping of $[h_1], [h_{2a}, h_{2b}, \ldots, h_{2e}], [h_{3a}, h_{3b}, \ldots, h_{3e}]$ versus the null grouping of $[h_1, h_{2a}, h_{2b}, \ldots, h_{2e}, h_{3a}, h_{3b}, \ldots, h_{3e}]$ could also be performed. Note that the generalized likelihood ratio test-statistic would have two degrees of freedom instead of one.

Measuring information

Even though likelihood ratio tests based on likelihoods computed directly for the observed data, which have captured the information loss due to uncertainty in phase and missing genotypes, can be relied on to give valid p-values, it would still be

of interest to know how much information had been lost due to the information being incomplete. Interestingly, one can measure information loss by considering a two-step procedure to evaluating statistical significance that appears natural but happens to be systematically anti-conservative. Suppose we calculate the maximum likelihood estimates for the population haplotype frequencies calculated under the alternative hypothesis that there are differences between the affected population and control population, and use these frequency estimates as estimates of the observed frequencies of haplotype counts in the affected sample and in the control sample. Suppose we then perform a likelihood ratio test treating these estimated haplotype counts as though they are the actual counts. We could also perform a Fisher's exact test, but we would then need to round off these estimated counts since they are in general non-integers. This test will in general be anti-conservative because treating the estimated counts as if they were exact counts ignores the uncertainty with the counts, overestimates the effective sample size and underestimates the sampling variation. It means that the chi-square likelihood-ratio test statistic calculated this way, denoted by Λ^*, will in general be bigger than Λ, the likelihood-ratio test-statistic calculated directly from the observed data as described in methods. But Λ^* is useful because the ratio Λ/Λ^* happens to be a good measure of information, or $1 - (\Lambda/\Lambda^*)$ is a measure of the fraction of information lost due to missing information. This information measure for haplotype analysis is described in Nicolae and Kong, Technical Report 537, Department of Statistics, University of Statistics, University of Chicago, Revised for *Biometrics* (2003) as a natural extension of information measures defined for linkage analysis, and is implemented in NEMO.

Statistical analysis.

For single marker association to the disease, the Fisher exact test can be used to calculate two-sided p-values for each individual allele. All p-values are presented unadjusted for multiple comparisons unless specifically indicated. The presented frequencies (for microsatellites, SNPs and haplotypes) are allelic frequencies as opposed to carrier frequencies. To minimize any bias due the relatedness of the patients who were recruited as families for the linkage analysis, first and second-degree relatives can be eliminated from the patient list. Furthermore, the test can be repeated for association correcting for any remaining relatedness among the patients, by extending a variance adjustment procedure described in Risch, N. & Teng, J. (*Genome Res.,* 8:1278-1288 (1998)). The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. (*ibid*) for sibships so that it can be applied to general familial relationships, and present both adjusted and unadjusted p-values for comparison. The differences are in general very small as expected. To assess the significance of single-marker association corrected for multiple testing we carried out a randomisation test using the same genotype data. Cohorts of patients and controls can be randomized and the association analysis redone multiple times (e.g., up to 500,000 times) and the p-value is the fraction of replications that produced a p-value for some marker allele that is lower than or equal to the p-value we observed using the original patient and control cohorts.

For both single-marker and haplotype analyses, relative risk (RR) and the population attributable risk (PAR) can be calculated assuming a multiplicative model (haplotype relative risk model), (Terwilliger, J.D. & Ott, J., *Hum Hered,* 42, 337-46 (1992) and Falk, C.T. & Rubinstein, P, *Ann Hum Genet* 51 (Pt 3), 227-33 (1987)), i.e., that the risks of the two alleles/haplotypes a person carries multiply. For example, if RR is the risk of A relative to a, then the risk of a person homozygote AA will be RR times that of a heterozygote Aa and RR^2 times that of a homozygote aa. The multiplicative model has a nice property that simplifies analysis and computations — haplotypes are independent, *i.e.,* in Hardy-Weinberg equilibrium, within the affected population as well as within the control population. As a

consequence, haplotype counts of the affecteds and controls each have multinomial distributions, but with different haplotype frequencies under the alternative hypothesis. Specifically, for two haplotypes h_i and h_j, risk(h_i)/risk(h_j) = $(f_i/p_i)/(f_j/p_j)$, where f and p denote respectively frequencies in the affected population and in the control population. While there is some power loss if the true model is not multiplicative, the loss tends to be mild except for extreme cases. Most importantly, p-values are always valid since they are computed with respect to null hypothesis.

In general, haplotype frequencies are estimated by maximum likelihood and tests of differences between cases and controls are performed using a generalized likelihood ratio test (Rice, J.A. *Mathematical Statistics and Data Analysis*, 602 (International Thomson Publishing, (1995)). deCODE's haplotype analysis program called NEMO, which stands for NEsted MOdels, can be used to calculate all the haplotype results. To handle uncertainties with phase and missing genotypes, it is emphasized that we do not use a common two-step approach to association tests, where haplotype counts are first estimated, possibly with the use of the EM algorithm, Dempster, (A.P., Laird, N.M. & Rubin, D.B., *Journal of the Royal Statistical Society B,* 39, 1-38 (1971)) and then tests are performed treating the estimated counts as though they are true counts, a method that can sometimes be problematic and may require randomisation to properly evaluate statistical significance. Instead, with NEMO, maximum likelihood estimates, likelihood ratios and p-values are computed with the aid of the EM-algorithm directly for the observed data, and hence the loss of information due to uncertainty with phase and missing genotypes is automatically captured by the likelihood ratios. Even so, it is of interest to know how much information is retained, or lost, due to incomplete information. Described herein is such a measure that is natural under the likelihood framework. For a fixed set of markers, the simplest tests performed compare one selected haplotype against all the others. Call the selected haplotype h_1 and the others $h_2, ..., h_k$. Let $p_1, \ldots, p_k$ denote the population frequencies of the haplotypes in the controls, and $f_1, \ldots, f_k$ denote the population frequencies of the haplotypes in the affecteds. Under the null hypothesis, $f_i = p_i$ for all i. The alternative model we use for the test assumes $h_2, ..., h_k$ to have the same risk while h_1 is allowed to have a different risk. This implies that while p_1 can be

different from f_1, $f_i/(f_2+\ldots+f_k) = p_i/(p_2+\ldots+p_k) = \beta_i$ for $i = 2, \ldots, k$. Denoting f_1/p_1 by r, and noting that $\beta_2+\ldots+\beta_k = 1$, the test statistic based on generalized likelihood ratios is

$$\Lambda = 2\left[\ell(\hat{r},\hat{p}_1,\hat{\beta}_2,\ldots,\hat{\beta}_{k-1}) - \ell(1,\tilde{p}_1,\tilde{\beta}_2,\ldots,\tilde{\beta}_{k-1})\right]$$

where ℓ denotes $\log_e$likelihood and $\tilde{}$ and ^ denote maximum likelihood estimates under the null hypothesis and alternative hypothesis respectively. Λ has asymptotically a chi-square distribution with 1-df, under the null hypothesis. Slightly more complicated null and alternative hypotheses can also be used. For example, let h_1 be G0, h_2 be GX and h_3 be AX. When comparing G0 against GX, *i.e.,* this is the test which gives estimated RR of 1.46 and p-value = 0.0002, the null assumes G0 and GX have the same risk but AX is allowed to have a different risk. The alternative hypothesis allows, for example, three haplotype groups to have different risks. This implies that, under the null hypothesis, there is a constraint that $f_1/p_1 = f_2/p_2$, or $w = [f_1/p_1]/[f_2/p_2] = 1$. The test statistic based on generalized likelihood ratios is

$$\Lambda = 2\left[\ell(\hat{p}_1,\hat{f}_1,\hat{p}_2,\hat{w}) - \ell(\tilde{p}_1,\tilde{f}_1,\tilde{p}_2,1)\right]$$

that again has asymptotically a chi-square distribution with 1-df under the null hypothesis. If there are composite haplotypes (for example, h_2 and h_3), that is handled in a natural manner under the nested models framework.

LD between pairs of SNPs can be calculated using the standard definition of D' and R^2 (Lewontin, R., Genetics 49, 49-67 (1964) and Hill, W.G. & Robertson, A. Theor. Appl. Genet. 22, 226-231 (1968)).Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood and deviation from linkage equilibrium is evaluated by a likelihood ratio test. The definitions of D' and R^2 are extended to include microsatellites by averaging over the values for all possible allele combination of the two markers weighted by the marginal allele probabilities. When plotting all marker combination to elucidate the LD structure in a particular region, we plot D' in the upper left corner and the p-value in the lower right corner. In the LD plots the markers can be plotted equidistant rather than according to their physical location, if desired.

Statistical Methods for Linkage Analysis

Multipoint, affected-only allele-sharing methods can be used in the analyses to assess evidence for linkage. Results, both the LOD-score and the non-parametric linkage (NPL) score, can obtained using the program Allegro (Gudbjartsson *et al.*, *Nat. Genet. 25:*12-3, 2000). Our baseline linkage analysis uses the S_{pairs} scoring function (Whittemore, A.S., Halpern, J. (1994), *Biometrics 50*:118-27; Kruglyak L, *et al.* (1996), *Am J Hum Genet 58*:1347-63), the exponential allele-sharing model (Kong, A. and Cox, N.J. (1997), *Am J Hum Genet 61*:1179-88) and a family weighting scheme that is halfway, on the log-scale, between weighting each affected pair equally and weighting each family equally. The information measure we use is part of the Allegro program output and the information value equals zero if the marker genotypes are completely uninformative and equals one if the genotypes determine the exact amount of allele sharing by decent among the affected relatives (Gretarsdottir *et al.*, *Am. J. Hom. Genet, 70*:593-603, (2002)). We computed the P-values two different ways and here report the less significant result. The first P-value can be computed on the basis of large sample theory; the distribution of $Z_{lr} = \sqrt{(2[\log_e(10)LOD])}$ approximates a standard normal variable under the null hypothesis of no linkage (Kong, A. and Cox, N.J. (1997), *Am J Hum Genet 61*:1179-88). The second P-value can be calculated by comparing the observed LOD-score with its complete data sampling distribution under the null hypothesis (e.g., Gudbjartsson *et al., Nat. Genet. 25:*12-3, 2000). When the data consist of more than a few families, these two P-values tend to be very similar.

KITS

Kits (*e.g.*, reagent kits) useful in the methods of diagnosis comprise components useful in any of the methods described herein, including for example, hybridization probes or primers as described herein (*e.g.*, labeled probes or primers), reagents for detection of labeled molecules, restriction enzymes (*e.g.*, for RFLP analysis), allele-specific oligonucleotides, antibodies which bind to altered or to non-altered (native) FLAP polypeptide, means for amplification of nucleic acids comprising a FLAP, or means for analyzing the nucleic acid sequence of a nucleic acid described herein, or for analyzing the amino acid sequence of a polypeptide as described herein, etc. In one embodiment, a kit for diagnosing susceptibility to MI or stroke can comprise primers for nucleic acid amplification of a region in the FLAP nucleic acid comprising an at-risk haplotype that is more frequently present in an individual having MI or stroke or susceptible to MI or stroke. The primers can be designed using portions of the nucleic acids flanking SNPs that are indicative of MI or stroke. In a particularly preferred embodiment, the primers are designed to amplify regions of the FLAP nucleic acid associated with an at-risk haplotype for MI or stroke, or more particularly the haplotypes defined by the following SNPs: DG00AAFIU, SG13S25, DG00AAJFF, DG00AAHII, DG00AAHID, B_SNP_310657, SG13S30, SG13S32, SG13S42, and SG13S35 , at the locus on chromosome 13q12. In other preferred embodiments, the primers are designed to amplify regions of the FLAP nucleic acid associated with a haplotype such as haplotype B4, B5, B6, A4, A5, HapB, a haplotype shown in Table 4, and/or a haplotype shown in Table 5, and/or a haplotype shown in Table 13.

DIAGNOSIS OF FLAP-RELATED DISEASE

Although the methods of diagnosis above have been described in the context of diagnosing susceptibility to MI or stroke, the methods can also be used to identify FLAP-associated MI and/or stroke. For example, individuals who have experienced MI and/or stroke can be assessed to determine whether the presence in the individual of a polymorphism in a FLAP nucleic acid, or the presence of an at-risk haplotype in

the individual, as described above, could have been a contributing factor to the MI and/or stroke. As used herein, the terms, "FLAP-associated MI" and "FLAP-associated stroke," refer to the occurrence of an MI or stroke in an individual who has a polymorphism in a FLAP nucleic acid or an at-risk FLAP haplotype. Identification of FLAP-associated MI or stroke facilitates treatment planning, as treatment can be designed and therapeutics selected to target components of the FLAP pathway.

In one embodiment of the invention, diagnosis of FLAP-associated MI or stroke, is made by detecting a polymorphism in a FLAP nucleic acid as described herein. A polymorphism in a FLAP nucleic acid is described above. A test sample of genomic DNA, RNA, or cDNA, is obtained from an individual who has had at least one MI and/or stroke, to determine whether the MI or stroke is FLAP-associated. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in a nucleic acid is present, and/or to determine which splicing variant(s) encoded by the FLAP is present. If the FLAP nucleic acid has the polymorphism, or is the splicing variant associated with disease, then the presence of the polymorphism or the splicing variant is indicative of FLAP-associated MI or stroke.

For example, in one embodiment, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used to detect the polymorphism. In other embodiments, mutation analysis by restriction digestion or sequence analysis can also be used, as can allele-specific oligonucleotides, or quantitative PCR (kinetic thermal cycling). Diagnosis of FLAP-associated MI or can also be made by examining expression and/or composition of a FLAP polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. A test sample from an individual is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by a FLAP nucleic acid, or for the presence of a particular variant encoded by a FLAP nucleic acid. An alteration in expression of a polypeptide encoded by a FLAP nucleic acid can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by a FLAP

nucleic acid is an alteration in the qualitative polypeptide expression (*e.g.*, expression of an altered FLAP polypeptide or of a different splicing variant.

In other embodiments, the invention pertains to a method for the diagnosis and identification of FLAP-associated myocardial infarction or stroke in an individual, by identifying the presence of an at-risk haplotype in FLAP as described in detail herein. For example, the haplotypes described herein in Tables 4 and 5, are found more frequently in individuals having MI, or stroke than in individuals not affected by these diseases. Therefore, these haplotypes have predictive value for detecting FLAP-associated MI or stroke in an individual. In certain embodiments, an at-risk haplotype is characterized by the presence of polymorphism(s) shown in Table 3. In other embodiments, the at-risk haplotype is selected from the group consisting of: haplotype B4, B5, B6, A4 and A5. The at-risk haplotype can also comprise a combination of the markers in the haplotypes B4, B5, B6, A4 and/or A5. In further embodiments, the at-risk haplotype can be haplotype HapB. The methods described herein can be used to assess a sample from an individual for the presence or absence of an at-risk haplotype; the presence of an at-risk haplotype is indicative of FLAP-associated MI or stroke.

In representative embodiments of the invention, a method of diagnosing FLAP-associated myocardial infarction or stroke in an individual who has had a myocardial infarction and/or a stroke, comprises detecting a polymorphism in a FLAP nucleic acid, wherein the presence of the polymorphism in the nucleic acid is indicative of FLAP-associated myocardial infarction or stroke. Alternatively, a method of diagnosing FLAP-associated myocardial infarction or stroke in an individual who has had a myocardial infarction and/or a stroke, comprises detecting an alteration in the expression or composition of a polypeptide encoded by a FLAP nucleic acid in a test sample, in comparison with the expression or composition of a polypeptide encoded by a FLAP nucleic acid in a control sample, wherein the presence of an alteration in expression or composition of the polypeptide in the test sample is indicative of FLAP-associated myocardial infarction or stroke. In addition, a method of diagnosing FLAP-associated myocardial infarction or stroke in an individual who has had a myocardial infarction and/or a stroke, comprises

determining the presence or absence in the individual of a haplotype selected from haplotypes shown in Table 4, haplotypes shown in Table 5, and haplotypes shown in Table 13, wherein the presence of the haplotype is diagnostic of FLAP-associated myocardial infarction or stroke. Also, a method of diagnosing FLAP-associated myocardial infarction or stroke in an individual who has had a myocardial infarction and/or a stroke, comprises determining the presence or absence in the individual of haplotype B4, B5, B6, A4, A5, or HapB, wherein the presence of the haplotype is diagnostic of FLAP-associated myocardial infarction or stroke.

SCREENING ASSAYS AND AGENTS IDENTIFIED THERBY

The invention provides methods (also referred to herein as "screening assays") for identifying the presence of a nucleotide that hybridizes to a nucleic acid of the invention, as well as for identifying the presence of a polypeptide encoded by a nucleic acid of the invention. In one embodiment, the presence (or absence) of a nucleic acid molecule of interest (*e.g.*, a nucleic acid that has significant homology with a nucleic acid of the invention) in a sample can be assessed by contacting the sample with a nucleic acid comprising a nucleic acid of the invention (*e.g.*, a nucleic acid having the sequence of one of SEQ ID NOs: 1or 3 or the complement thereof, or a nucleic acid encoding an amino acid having the sequence of SEQ ID NO: 2, or a fragment or variant of such nucleic acids), under stringent conditions as described above, and then assessing the sample for the presence (or absence) of hybridization. In a preferred embodiment, high stringency conditions are conditions appropriate for selective hybridization. In another embodiment, a sample containing a nucleic acid molecule of interest is contacted with a nucleic acid containing a contiguous nucleic acid sequence (*e.g.*, a primer or a probe as described above) that is at least partially complementary to a part of the nucleic acid molecule of interest (*e.g.*, a FLAP nucleic acid), and the contacted sample is assessed for the presence or absence of hybridization. In a preferred embodiment, the nucleic acid containing a contiguous nucleic acid sequence is completely complementary to a part of the nucleic acid molecule of interest.

In any of these embodiments, all or a portion of the nucleic acid of interest can be subjected to amplification prior to performing the hybridization.

In another embodiment, the presence (or absence) of a polypeptide of interest, such as a polypeptide of the invention or a fragment or variant thereof, in a sample can be assessed by contacting the sample with an antibody that specifically hybridizes to the polypeptide of interest (*e.g.*, an antibody such as those described above), and then assessing the sample for the presence (or absence) of binding of the antibody to the polypeptide of interest.

In another embodiment, the invention provides methods for identifying agents (*e.g.*, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes which alter (*e.g.*, increase or decrease) the activity of the polypeptides described herein, or which otherwise interact with the polypeptides herein. For example, such agents can be agents which bind to polypeptides described herein (*e.g.*, binding agent for members of the leukotriene pathway, such as FLAP binding agents); which have a stimulatory or inhibitory effect on, for example, activity of polypeptides of the invention; or which change (*e.g.*, enhance or inhibit) the ability of the polypeptides of the invention to interact with members of the leukotriene pathway binding agents (*e.g.*, receptors or other binding agents); or which alter posttranslational processing of the leukotriene pathway member polypeptide, such as a FLAP polypeptide (*e.g.*, agents that alter proteolytic processing to direct the polypeptide from where it is normally synthesized to another location in the cell, such as the cell surface; agents that alter proteolytic processing such that more polypeptide is released from the cell, etc.)

In one embodiment, the invention provides assays for screening candidate or test agents that bind to or modulate the activity of polypeptides described herein (or biologically active portion(s) thereof), as well as agents identifiable by the assays. Test agents can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is

limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S., *Anticancer Drug Des.* 12:145 (1997)).

In one embodiment, to identify agents which alter the activity of a FLAP polypeptide, a cell, cell lysate, or solution containing or expressing a FLAP polypeptide (*e.g.*, SEQ ID NO: 2 or another splicing variant encoded by a FLAP nucleic acid, such as a nucleic acid comprising a SNP as shown in Table 3), or a fragment or derivative thereof (as described above), can be contacted with an agent to be tested; alternatively, the polypeptide can be contacted directly with the agent to be tested. The level (amount) of FLAP activity is assessed (*e.g.*, the level (amount) of FLAP activity is measured, either directly or indirectly), and is compared with the level of activity in a control (*i.e.*, the level of activity of the FLAP polypeptide or active fragment or derivative thereof in the absence of the agent to be tested). If the level of the activity in the presence of the agent differs, by an amount that is statistically significant, from the level of the activity in the absence of the agent, then the agent is an agent that alters the activity of a FLAP polypeptide. An increase in the level of FLAP activity in the presence of the agent relative to the activity in the absence of the agent, indicates that the agent is an agent that enhances (is an agonist of) FLAP activity. Similarly, a decrease in the level of FLAP activity in the presence of the agent, relative to the activity in the absence of the agent, indicates that the agent is an agent that inhibits (is an antagonist of) FLAP activity. In another embodiment, the level of activity of a FLAP polypeptide or derivative or fragment thereof in the presence of the agent to be tested, is compared with a control level that has previously been established. A statistically significant difference in the level of the activity in the presence of the agent from the control level indicates that the agent alters FLAP activity.

The present invention also relates to an assay for identifying agents which alter the expression of a FLAP nucleic acid (*e.g.*, antisense nucleic acids, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes; which alter (*e.g.*, increase or decrease) expression (*e.g.*, transcription or translation) of the nucleic acid or which

otherwise interact with the nucleic acids described herein, as well as agents identifiable by the assays. For example, a solution containing a nucleic acid encoding a FLAP polypeptide (*e.g.*, a FLAP nucleic acid) can be contacted with an agent to be tested. The solution can comprise, for example, cells containing the nucleic acid or cell lysate containing the nucleic acid; alternatively, the solution can be another solution that comprises elements necessary for transcription/translation of the nucleic acid. Cells not suspended in solution can also be employed, if desired. The level and/or pattern of FLAP expression (*e.g.*, the level and/or pattern of mRNA or of protein expressed, such as the level and/or pattern of different splicing variants) is assessed, and is compared with the level and/or pattern of expression in a control (*i.e.*, the level and/or pattern of the FLAP expression in the absence of the agent to be tested). If the level and/or pattern in the presence of the agent differ, by an amount or in a manner that is statistically significant, from the level and/or pattern in the absence of the agent, then the agent is an agent that alters the expression of the FLAP nucleic acid. Enhancement of FLAP expression indicates that the agent is an agonist of FLAP activity. Similarly, inhibition of FLAP expression indicates that the agent is an antagonist of FLAP activity.

In another embodiment, the level and/or pattern of FLAP polypeptide(s) (*e.g.*, different splicing variants) in the presence of the agent to be tested, is compared with a control level and/or pattern that have previously been established. A level and/or pattern in the presence of the agent that differs from the control level and/or pattern by an amount or in a manner that is statistically significant indicates that the agent alters FLAP expression.

In another embodiment of the invention, agents which alter the expression of a FLAP nucleic acid or which otherwise interact with the nucleic acids described herein, can be identified using a cell, cell lysate, or solution containing a nucleic acid encoding the promoter region of the FLAP nucleic acid operably linked to a reporter gene. After contact with an agent to be tested, the level of expression of the reporter gene (*e.g.*, the level of mRNA or of protein expressed) is assessed, and is compared with the level of expression in a control (*i.e.*, the level of the expression of the reporter gene in the absence of the agent to be tested). If the level in the presence of

the agent differs, by an amount or in a manner that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters the expression of the FLAP nucleic acid, as indicated by its ability to alter expression of a nucleic acid that is operably linked to the FLAP nucleic acid promoter.

Enhancement of the expression of the reporter indicates that the agent is an agonist of FLAP activity. Similarly, inhibition of the expression of the reporter indicates that the agent is an antagonist of FLAP activity. In another embodiment, the level of expression of the reporter in the presence of the test agent, is compared with a control level that has previously been established. A level in the presence of the agent that differs from the control level by an amount or in a manner that is statistically significant indicates that the agent alters expression.

Agents which alter the amounts of different splicing variants encoded by a FLAP nucleic acid (*e.g.*, an agent which enhances activity of a first splicing variant, and which inhibits activity of a second splicing variant), as well as agents which are agonists of activity of a first splicing variant and antagonists of activity of a second splicing variant, can easily be identified using these methods described above.

In other embodiments of the invention, assays can be used to assess the impact of a test agent on the activity of a polypeptide relative to a FLAP binding agent. For example, a cell that expresses a compound that interacts with a FLAP nucleic acid (herein referred to as a "FLAP binding agent", which can be a polypeptide or other molecule that interacts with a FLAP nucleic acid, such as a receptor, or another molecule, such as 5-LO) is contacted with a FLAP in the presence of a test agent, and the ability of the test agent to alter the interaction between the FLAP and the FLAP binding agent is determined. Alternatively, a cell lysate or a solution containing the FLAP binding agent, can be used. An agent which binds to the FLAP or the FLAP binding agent can alter the interaction by interfering with, or enhancing the ability of the FLAP to bind to, associate with, or otherwise interact with the FLAP binding agent. Determining the ability of the test agent to bind to a FLAP nucleic acid or a FLAP nucleic acid binding agent can be accomplished, for example, by coupling the test agent with a radioisotope or enzymatic label such that binding of the test agent to the polypeptide can be determined by detecting the labeled with ^{125}I, ^{35}S, ^{14}C or ^{3}H,

either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test agents can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. It is also within the scope of this invention to determine the ability of a test agent to interact with the polypeptide without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test agent with a FLAP or a FLAP binding agent without the labeling of either the test agent, FLAP, or the FLAP binding agent. McConnell, H.M. *et al.*, *Science* 257:1906-1912 (1992). As used herein, a "microphysiometer" (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and polypeptide.

Thus, these receptors can be used to screen for compounds that are agonists for use in treating a disease or condition associated with FLAP or a susceptibility to a disease or condition associated with FLAP, or antagonists for studying a susceptibility to a disease or condition associated with FLAP (*e.g.*, MI or stroke). Drugs can be designed to regulate FLAP activation, that in turn can be used to regulate signaling pathways and transcription events of genes downstream or of proteins or polypeptides interacting with FLAP (*e.g.*, 5-LO).

In another embodiment of the invention, assays can be used to identify polypeptides that interact with one or more FLAP polypeptides, as described herein. For example, a yeast two-hybrid system such as that described by Fields and Song (Fields, S. and Song, O., *Nature* 340:245-246 (1989)) can be used to identify polypeptides that interact with one or more FLAP polypeptides. In such a yeast two-hybrid system, vectors are constructed based on the flexibility of a transcription factor that has two functional domains (a DNA binding domain and a transcription activation domain). If the two domains are separated but fused to two different proteins that interact with one another, transcriptional activation can be achieved, and transcription of specific markers (*e.g.*, nutritional markers such as His and Ade, or

color markers such as lacZ) can be used to identify the presence of interaction and transcriptional activation. For example, in the methods of the invention, a first vector is used which includes a nucleic acid encoding a DNA binding domain and also a FLAP polypeptide, splicing variant, or fragment or derivative thereof, and a second vector is used which includes a nucleic acid encoding a transcription activation domain and also a nucleic acid encoding a polypeptide which potentially may interact with the FLAP polypeptide, splicing variant, or fragment or derivative thereof (*e.g.*, a FLAP polypeptide binding agent or receptor). Incubation of yeast containing the first vector and the second vector under appropriate conditions (*e.g.*, mating conditions such as used in the Matchmaker™ system from Clontech (Palo Alto, California, USA)) allows identification of colonies that express the markers of interest. These colonies can be examined to identify the polypeptide(s) that interact with the FLAP polypeptide or fragment or derivative thereof. Such polypeptides may be useful as agents that alter the activity of expression of a FLAP polypeptide, as described above.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the FLAP, the FLAP binding agent, or other components of the assay on a solid support, in order to facilitate separation of complexed from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. Binding of a test agent to the polypeptide, or interaction of the polypeptide with a binding agent in the presence and absence of a test agent, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein (*e.g.*, a glutathione-S-transferase fusion protein) can be provided which adds a domain that allows a FLAP nucleic acid or a FLAP binding agent to be bound to a matrix or other solid support.

In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell, cell lysate, or solution containing a nucleic acid encoding a FLAP nucleic acid is contacted with a test agent and the expression of appropriate mRNA or polypeptide (*e.g.*, splicing variant(s)) in the cell, cell lysate, or solution, is determined. The level of expression of appropriate mRNA or polypeptide(s) in the presence of the test agent is compared to the level of

expression of mRNA or polypeptide(s) in the absence of the test agent. The test agent can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or polypeptide is greater (statistically significantly greater) in the presence of the test agent than in its absence, the test agent is identified as a stimulator or enhancer of the mRNA or polypeptide expression. Alternatively, when expression of the mRNA or polypeptide is less (statistically significantly less) in the presence of the test agent than in its absence, the test agent is identified as an inhibitor of the mRNA or polypeptide expression. The level of mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting mRNA or polypeptide.

In yet another embodiment, the invention provides methods for identifying agents (*e.g.,* fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) which alter (*e.g.,* increase or decrease) the activity of a member of leukotriene pathway binding agent, such as a FLAP binding agent (*e.g.*, 5-LO), as described herein. For example, such agents can be agents which have a stimulatory or inhibitory effect on, for example, the activity of a member of leukotriene pathway binding agent, such as a FLAP binding agent; which change (*e.g.,* enhance or inhibit) the ability a member of leukotriene pathway binding agents, (*e.g.,* receptors or other binding agents) to interact with the polypeptides of the invention; or which alter posttranslational processing of the member of leukotriene pathway binding agent, (*e.g.,* agents that alter proteolytic processing to direct the member of the leukotriene pathway binding agent from where it is normally synthesized to another location in the cell, such as the cell surface; agents that alter proteolytic processing such that more active binding agent is released from the cell, etc.).

For example, the invention provides assays for screening candidate or test agents that bind to or modulate the activity of a member of the leukotriene pathway (or enzymatically active portion(s) thereof), as well as agents identifiable by the assays. As described above, test agents can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries;

synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. *Anticancer Drug Des., 12*:145 (1997)). In one embodiment, to identify agents which alter the activity of a member of the leukotriene pathway (such as a FLAP binding agent), a cell, cell lysate, or solution containing or expressing a binding agent (*e.g.,* 5-LO, or a leukotriene pathway member receptor), or a fragment (*e.g.,* an enzymatically active fragment) or derivative thereof, can be contacted with an agent to be tested; alternatively, the binding agent (or fragment or derivative thereof) can be contacted directly with the agent to be tested. The level (amount) of binding agent activity is assessed (either directly or indirectly), and is compared with the level of activity in a control (*i.e.*, the level of activity in the absence of the agent to be tested). If the level of the activity in the presence of the agent differs, by an amount that is statistically significant, from the level of the activity in the absence of the agent, then the agent is an agent that alters the activity of the member of the leukotriene pathway. An increase in the level of the activity relative to a control, indicates that the agent is an agent that enhances (is an agonist of) the activity. Similarly, a decrease in the level of activity relative to a control, indicates that the agent is an agent that inhibits (is an antagonist of) the activity. In another embodiment, the level of activity in the presence of the agent to be tested, is compared with a control level that has previously been established. A level of the activity in the presence of the agent that differs from the control level by an amount that is statistically significant indicates that the agent alters the activity.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a test agent that is a modulating agent, an antisense nucleic acid molecule, a specific antibody, or a polypeptide-binding agent) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent

identified as described herein can be used in an animal model to determine the mechanism of action of such an agent.

Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein. In addition, an agent identified as described herein can be used to alter activity of a polypeptide encoded by a FLAP nucleic acid, or to alter expression of a FLAP nucleic acid, by contacting the polypeptide or the nucleic acid (or contacting a cell comprising the polypeptide or the nucleic acid) with the agent identified as described herein.

The present invention is now illustrated by the following Examples, which are not intended to be limiting in any way.

EXAMPLE 1 CORRELATION BETWEEN MI AND FLAP

SUBJECTS AND METHODS

Study population

Patients entering the study were defined from an infarction registry that includes all MIs (over 8,000 patients) in Iceland 1981-2000. This registry is a part of the World Health Organization MONICA Project (The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin. Epidemiol.* 1988; 41:105-14). Diagnosis of all patients in the registry follow strict diagnostic rules based on symptoms, electrocardiograms, cardiac enzymes, and necropsy findings.

Blood samples from 1342 MI patients, both cases with a family history and sporadic cases were collected. For each patient that participated, blood was collected from 2 relatives (unaffected or affected). Their genotypes were used to help with construction of haplotypes. In addition, blood samples from 624 unrelated controls were collected.

Linkage analysis

Extended families (pedigrees) by clustering related female MI patients were constructed into families such that each patient is related to at least one other patient within and including six meiotic events. The information regarding the relatedness of patients was obtained from an encrypted genealogy database that covers the entire Icelandic nation (Gulcher *et al., Eur. J. Hum. Genet.* 8: 739-742 (2000)). A genomewide scan was performed using a framework map of 1000 microsatellite markers, using protocols described elsewhere (Gretarsdottir S., *et al. Am. J. Hum. Genet.*, 70: 593-603, 2002)). The marker order and positions where obtained from a modified version of the Marshfield genetic map (Center for Medical Genetics, Marshfield Medical Research Foundation), using genetic mapping based on our own data, and from deCODE genetic's high resolution genetic map (Kong A., *et al., Nat. genet.,* 31: 241-247 (2002)). The population-based allele frequencies were constructed from a cohort of more than 30,000 Icelanders who have participated in genetic studies of various disease projects. Additional markers were genotyped within the locus on chromosome 13 to increase the information on identity by descent within the families. For those markers at least 180 Icelandic controls were genotyped to derive the population allele frequencies.

For statistical analysis, multipoint, affected only allele-sharing methods were used to assess evidence for linkage. All results, both the LOD and the non-parametric linkage (NPL) score, were obtained using the program ALLEGRO (Gudbjartsson D.F., *et al., Nat Genet.,* 25: 12-13(2000)). The baseline linkage analysis (Gretarsdottir S., *et al., Am. J. Hum. Genet.* 70: 593-603, (2002)) uses the Spairs scoring function (Whittermore AS, and Haplern J A., *Biometrics* 50: 118-127 (1994)) and Kruglyak *et al., Am. J. Hum. Genet.,* 58:1347-1363 (1996)) the exponential allele-sharing model (Kong A., and Cox N.J., *Am. J. Hum. Genet.* 61:1179-1188 (1997)), and a family weighting scheme which is halfway, on the log-scale, between weighing each affected pairs equally and weighing each family equally.

Ultra-fine mapping and haplotype analysis

A candidate susceptibility locus was defined as the region under the LOD score curve where the score was one lower than the highest lod score ((peak lod score −1)\one lod drop). This region (approx. 12Mb) was ultra-finemapped with microsatellite markers with an average spacing between markers of less than 100kb. All usable microsatellite markers that found in public databases and mapped within that region were used. In addition, microsatellite markers identified within the deCODE genetics sequence assembly of the human genome were used.

Haplotype analysis

The frequencies of haplotypes in the patient and the control groups using an expectation-maximization algorithm were estimated (Dempster A.P. *et al., J. R. Stat. Soc. B.* 39: 1-389 (1977)). An implementation of this algorithm that can handle missing genotypes and uncertainty with the phase was used. Under the null hypothesis, the patients and the controls are assumed to have identical frequencies. Using a likelihood approach, an alternative hypothesis where a candidate at-risk-haplotype is allowed to have a higher frequency in patients than controls, while the ratios of the frequencies of other haplotypes are assumed to be the same in both groups was tested. Likelihoods are maximized separately under both hypothesis and a corresponding 1-df likelihood ratio statistics is used to evaluate the statistic significance.

To look for at-risk-haplotypes in the 1-lod drop, association of all possible combinations of genotyped markers was studied, provided those markers spanned a region of size less than 1000 kb. Due to a certain amount of testing, the *p*-values were adjusted using simulations. The combined patient and control groups were randomly divided into two sets, equal in size to the original group of patients and controls. The haplotype analysis was then repeated and the most significant *p*-value registered was observed. This randomisation scheme was repeated over 100 times to construct an empirical distribution of *p*-values.

SNP haplotype association to MI

In an effort to identify SNP haplotypes that associate with MI we have typed SNPs identified mainly by sequencing the FLAP gene and the region flanking the gene. We genotyped a total number of 45 SNPs in 1343 patients and 624 unrelated controls. The largest subset of unrelated patients (related no closer than 4 meioses) was 921. We observed two correlated series of SNP haplotypes in excess in patients, denoted as A and B in Table 7. The length of the haplotypes varies between 33 and 69 Kb and cover one or two blocks of linkage disequilibrium. Both series of haplotypes contain the common allele 2 of the SNP SG13S25. All haplotypes in the A series contain the SNP DG00AAHID, while all haplotypes in the B series contain the SNP DG00AAHII. In the B series, the haplotypes B4, B5, and B6 have a relative risk (RR) greater than 2 and allelic frequencies above 10% (Table 1). The haplotypes in the A series have slightly lower RR and p-values, but higher allelic frequency (15-16%), and as such we also consider them interesting. The haplotypes in series B and A are strongly correlated, *i.e.* the B haplotypes define a subset of the A haplotypes. Hence, B haplotypes are more specific than A haplotypes. However, A haplotypes are more sensitive, *i.e.* they capture more individuals with the putative mutation, as is observed in the population attributable risk which is less for B than for A. Furthermore, these haplotypes show similar risk ratios and allelic frequency for early-onset patients (defined as onset of first MI before the age of 55). In addition, analyzing various groups of patients with known risk factors, such as hypertension, high cholesterol, smoking and diabetes, did not reveal any significant correlation with these haplotypes.

In conclusion, we have identified a series of correlated MI disease risk haplotypes, consisting of 4-6 SNPs, with relative risk greater than 2 and allelic frequency in MI patients greater than 10%. The length of the haplotypes varies between 39-68 kb. These haplotypes are carried by 19% (B5) to 29% (A4) of MI patients. Our results suggest that the 'at risk' haplotypes in the FLAP gene represent a new major independent risk factor for MI.

Discussion

In a genome wide search for susceptibility nucleic acids for MI, a locus to 13q12 was mapped. This locus was ultra-fine mapped with microsatellite markers. Haplotype analysis strongly suggested a nucleic acid for FLAP (ALOX5AP), as a susceptibility gene for MI.

The FLAP gene encodes for a protein that is required for leukotriene synthesis (LTA4, LTB4, LTC4, LTD4). Inhibitors of its function impede translocation of 5-lipoxygenase from the cytoplasm to the cell membrane and inhibit activation of 5-lipoxygenase. The leukotrienes are potent inflammatory lipid mediators derived from arachidonic acid that can potentially contribute to development of atherosclerosis and destabilization of atherosclerotic plaques through lipid oxidation and/or proinflammatory effects. Allen *et al.*, (*Circulation.* 97: 2406-2413(1998)) described a novel mechanism in which atherosclerosis is associated with the appearance of a leukotriene receptor(s) capable of inducing hyperreactivity of human epicardial coronary arteries in response to LTC4 and LTD4. Allen *et al.* show a photomicrograph of a section of human atherosclerotic coronary artery a positive staining of a number of members of the leukotriene pathway, including FLAP. Mehrabian *et al.* described the identification of 5-Lipoxygenase (5-LO) as a major gene contributing to atherosclerosis susceptibility in mice. Mehrabian *et al.* described that heterozygous deficiency for the enzyme in a knockout model decreased the atherosclerotic lesion size in LDL-/- mice by about 95%. Mehrabian *et al* show that the enzyme is expressed abundantly in macrophage-rich regions of atherosclerotic lesions, and suggested that 5-LO and/or its products might act locally to promote lesion development (Mehrabian *et al.*, *Circulation Research.* 91:120 (2002)).

Studies of FLAP inhibition in animal models of atheroscerosis are scarce. However, in a rabbit model of acute MI assesssed 72 hours after coronary artery ligation the FLAP-inhibitor BAYx1005 madedly reduced mortality, from 65% to 25%, and blocked the increase in CPK and neutrophil accumulation as well as the ECG-changes observed in sham treated animals (*J. Pharmacol. Exp. Ther.*, 276:332 (1996)).

Mutations and /or polymorphisms within the FLAP nucleic acid, and other

members of the same pathway (*i.e.,* 5-lipoxygenase, LTA4, LTB4, LTC4, and CysLT2 receptor), that show association with the disease, can be used as a diagnostic test. The members of the 5-LO pathway in particular are valuable therapeutic targets for myocardial infarction.

Table 1: The marker map for chromosome 13 used in the linkage analysis.

Location (cM)	Marker	Location (cM)	Marker
6	D13S175	63.9	D13S170
9.8	D13S1243	68.7	D13S265
13.5	D13S1304	73	D13S167
17.2	D13S217	76.3	D13S1241
21.5	D13S289	79.5	D13S1298
25.1	D13S171	81.6	D13S1267
28.9	D13S219	84.7	D13S1256
32.9	D13S218	85.1	D13S158
38.3	D13S263	87	D13S274
42.8	D13S326	93.5	D13S173
45.6	D13S153	96.7	D13S778
49.4	D13S1320	102.7	D13S1315
52.6	D13S1296	110.6	D13S285
55.9	D13S156	115	D13S293
59.8	D13S1306		

Table 2: Marker Map for the second step of Linkage Analysis

Location (cM)	Marker	Location (cM)	Marker
1.758	D13S175	42.585	D13S1248
9.235	D13S787	44.288	D13S1233
11.565	D13S1243	44.377	D13S263
16.898	D13S221	45.535	D13S325
17.454	D13S1304	45.536	D13S1270
18.011	D13S1254	45.537	D13S1276
18.59	D13S625	49.149	D13S326
19.308	D13S1244	49.532	D13S1272
19.768	D13S243	52.421	D13S168
22.234	D13S1250	52.674	D13S287
22.642	D13S1242	60.536	D13S1320
22.879	D13S217	64.272	D13S1296
25.013	D13S1299	71.287	D13S156
28.136	D13S289	76.828	D13S1306
28.678	D13S290	77.86	D13S170
29.134	D13S1287	82.828	D13S265
30.073	D13S260	91.199	D13S1241
31.98	D13S171	93.863	D13S1298
32.859	D13S267	97.735	D13S779
33.069	D13S1293	100.547	D13S1256
33.07	D13S620	102.277	D13S274
34.131	D13S220	111.885	D13S173
36.427	D13S219	112.198	D13S796
39.458	D13S1808	115.619	D13S778
40.441	D13S218	119.036	D13S1315
41.113	D13S1288	126.898	D13S285
41.996	D13S1253	131.962	D13S293

Table 3 shows the exons with positions that encode the FLAP protein, markers and SNPs identified within the genomic sequence by the methods described herein.

Table 3

NCBI build34; start on chr. 13 (bp)	NCBI build34; stop on chr. 13 (bp)	SNP/marker/ exon name	alias1	alias2	public SNP	Variation
28932432	28932432	SG13S421		DG00AAFQR	rs1556428	A/G
28960356	28960356	SG13S417		SNP13B_R1028729	rs1028729	C/T
28965803	28965803	SG13S418		SNP13B_Y1323898	rs1323898	A/G
28974627	28974627	SG13S44				A/G
28975101	28975101	SG13S45				C/G
28975315	28975315	SG13S46				A/G
28975353	28975353	SG13S50				C/T
28975774	28975774	SG13S52				A/G
28985244	28985244	SG13S53			rs1408167	A/C
28985303	28985303	SG13S55			rs1408169	A/G
28985423	28985423	SG13S56				G/T
28985734	28985734	SG13S57			rs6490471	C/T
28985902	28985902	SG13S58			rs6490472	A/G
29003869	29003869	SG13S59				C/G
29004696	29004696	SG13S60				A/G
29007670	29007670	SG13S419		SNP13B_K912392	rs912392	C/T
29015410	29015410	SG13S61				C/T
29025792	29025792	SG13S62				C/T
29026202	29026202	SG13S63			rs7997114	A/G
29026668	29026668	SG13S64				A/G
29038707	29038707	SG13S65				A/G
29042180	29042180	SG13S420		DG00AAFIV	rs2248564	A/T
29049355	29049355	SG13S66				A/G
29049446	29049446	SG13S67				C/T
29050416	29050416	SG13S69				A/C
29059348	29059348	SG13S70				A/G
29059383	29059383	SG13S71				A/G
29059402	29059402	SG13S72				G/T
29063702	29063949	D13S289				
29064359	29064753	DG13S166				
29066272	29066272	SG13S73				A/G
29070551	29070551	SG13S99	SNP_13_Y1323892	DG00AAFIU	rs1323892	C/T
29081983	29081983	SG13S382	FLA267479			A/G
29082200	29082200	SG13S383	FLA267696			A/G
29082357	29082357	SG13S384	FLA267853			A/G
29083350	29083350	SG13S381	FLA268846	DG00AAJER		C/G
29083518	29083518	SG13S366	FLA269014	DG00AAJES	rs4312166	A/G
29085102	29085102	SG13S385	FLA270742			C/T
29085190	29085190	SG13S386	FLA270830			A/G
29086224	29086224	SG13S1	FLA271864			G/T
29087473	29087473	SG13S2	FLA273371			A/G
29088090	29088090	SG13S367	FLA273988	DG00AAJEU	rs4474551	A/G
29088186	29088186	SG13S388	FLA274084			A/G
29088473	29088473	SG13S10	FLA274371			A/T
29089044	29089044	SG13S3	FLA274942			C/T
29089886	29089886	SG13S368	FLA275784	DG00AAJEV		C/T
29090025	29090025	SG13S369	FLA275923	DG00AAJEW		G/T
29090054	29090054	SG13S370	FLA275952	DG00AAJEX		A/G
29090997	29090997	SG13S4	FLA276895			G/C
29091307	29091307	SG13S5	FLA277205		rs4238133	G/T
29091580	29091580	SG13S389	FLA277478			A/G

29091780	29091780	SG13S90	FLA277678			A/C
29092287	29092287	SG13S390	FLA278185		rs5004913	A/G
29092536	29092536	SG13S6	FLA278434			A/G
29092594	29092594	SG13S391	FLA278492			A/G
29092947	29092947	SG13S392	FLA278845			G/T
29093964	29093964	SG13S371	FLA279888	DG00AAJEY	rs4409939	A/G
29094259	29094259	SG13S372	FLA280183	DG00AAJEZ		A/G
29094999	29094999	SG13S393	FLA280923			A/T
29096688	29096688	SG13S373	FLA282612	DG00AAJFA		A/G
29096813	29096813	SG13S374	FLA282737	DG00AAJFB		A/G
29096874	29096874	SG13S375	FLA282798	DG00AAJFC		C/T
29096962	29096962	SG13S376	FLA282886	DG00AAJFD		A/G
29097476	29097476	SG13S394	FLA283400			C/G
29097553	29097553	SG13S25	FLA283477			A/G
29098486	29098486	SG13S395	FLA284410			A/G
29098891	29098891	SG13S396	FLA284815			A/C
29098979	29098979	SG13S397	FLA284903			C/T
29101965	29101965	SG13S377	FLA287889	DG00AAJFF		A/G
29103909	29103909	SG13S189	FLA289833			C/G
29104271	29104271	SG13S100	FLA290195	DG00AAHIK	rs4073259	A/G
29104629	29104629	SG13S398	FLA290553			C/G
29104646	29104646	SG13S94	FLA290570		rs4073261	C/T
29105099	29105099	SG13S101	FLA291023		rs4075474	C/T
29106329	29106329	SG13S95	FLA292253			G/T
29106652	29106652	SG13S102	FLA292576			A/T
29107138	29107138	SG13S103	FLA293062			C/T
29107404	29107404	SG13S104	FLA293328			A/G
29107668	29107812	EXON1				
29107830	29107830	SG13S191	FLA293754	DG00AAFJT	rs4769055	A/C
29108398	29108398	SG13S105	FLA294322			A/G
29108579	29108579	SG13S106	FLA294503	DG00AAHII		A/G
29108919	29108919	SG13S107	FLA294843		rs4075131	A/G
29108972	29108972	SG13S108	FLA294896		rs4075132	C/T
29109112	29109112	SG13S109	FLA295036			A/G
29109182	29109182	SG13S110	FLA295106			A/G
29109344	29109344	SG13S111	FLA295268		rs4597169	C/T
29109557	29109557	SG13S112	FLA295481			C/T
29109773	29109773	SG13S113	FLA295697		rs4293222	C/G
29110096	29110096	SG13S114	FLA296020	DG00AAHID		A/T
29110178	29110178	SG13S115	FLA296102			A/T
29110508	29110508	SG13S116	FLA296432		rs4769871	C/T
29110630	29110630	SG13S117	FLA296554		rs4769872	A/G
29110689	29110689	SG13S118	FLA296613		rs4769873	C/T
29110862	29110862	SG13S119	FLA296786			A/G
29111889	29111889	SG13S120	FLA297813			C/T
29112174	29112174	SG13S121	FLA298098	DG00AAHIJ	rs4503649	A/G
29112264	29112264	SG13S122	FLA298188	DG00AAHIH		A/G
29112306	29112306	SG13S123	FLA298230			C/T
29112455	29112455	SG13S43	FLA298379		rs3885907	A/C
29112583	29112583	SG13S399	FLA298507			A/C
29112680	29112680	SG13S124	FLA298604		rs3922435	C/T
29113139	29113139	SG13S125	FLA299063			A/G
29114056	29114056	SG13S400	FLA299980			A/G
29114738	29114738	SG13S126	FLA300662			A/G
29114940	29114940	SG13S127	FLA300864			A/G
29115878	29115878	SG13S128	FLA302094		rs4254165	A/G

29116020	29116020	SG13S129	FLA302236		rs4360791	A/G
29116068	29116068	SG13S130	FLA302284			G/T
29116196	29116296	EXON2				
29116249	29116249	SG13S190	FLA302465			C/T
29116308	29116308	SG13S192	FLA302524	B_SNP_302524	rs3803277	A/C
29116344	29116344	SG13S193	FLA302560			A/G
29116401	29116401	SG13S88	FLA302617	B_SNP_302617	rs3803278	C/T
29116688	29116688	SG13S131	FLA302904			C/T
29117133	29117133	SG13S132	FLA303349			A/C
29117546	29117546	SG13S133	FLA303762		rs4356336	C/T
29117553	29117553	SG13S38	FLA303769		rs4584668	A/T
29117580	29117580	SG13S134	FLA303796			C/T
29117741	29117741	SG13S135	FLA303957		rs4238137	C/T
29117954	29117954	SG13S136	FLA304170		rs4147063	C/T
29118118	29118118	SG13S137	FLA304334	DG00AAHIG	rs4147064	C/T
29118815	29118815	SG13S86	FLA305031			A/G
29118873	29118873	SG13S87	FLA305089	DG00AAHOJ		A/G
29119069	29119069	SG13S138	FLA305285			C/T
29119138	29119138	SG13S139	FLA305354			C/G
29119289	29119289	SG13S140	FLA305505			A/G/T
29119462	29119462	SG13S141	FLA305678			C/T
29119740	29119740	SG13S39	FLA305956			G/T
29120939	29120939	SG13S142	FLA307155		rs4387455	C/T
29120949	29120949	SG13S143	FLA307165		rs4254166	C/T
29121342	29121342	SG13S144	FLA307558		rs4075692	A/G
29121572	29121572	SG13S145	FLA307788			C/G
29121988	29121988	SG13S146	FLA308204			C/T
29122253	29122253	SG13S26	FLA308469			C/T
29122283	29122283	SG13S27	FLA308499			A/G
29122294	29122294	SG13S147	FLA308510			C/T
29122298	29122298	SG13S28	FLA308514			G/T
29122311	29122311	SG13S148	FLA308527			G/T
29123370	29123370	SG13S98	FLA309586			G/T
29123635	29123635	SG13S149	FLA309851			A/G
29123643	29123643	SG13S29	FLA309859			A/C
29124188	29124259	EXON3				
29124441	29124441	SG13S89	FLA310657	B_SNP_310657	rs4769874	A/G
29124906	29124906	SG13S96	FLA311122		rs4072653	A/G
29125032	29125032	SG13S150	FLA311248			C/G
29125521	29125521	SG13S401	FLA311737			C/T
29125822	29125822	SG13S151	FLA312038			C/T
29125840	29125840	SG13S30	FLA312056			G/T
29127301	29127301	SG13S31	FLA313550			C/T
29128080	29128162	EXON4				
29128284	29128284	SG13S152	FLA314500			C/G
29128316	29128316	SG13S402	FLA314532		rs4468448	C/T
29128798	29128798	SG13S403	FLA315014		rs4399410	A/G
29129016	29129016	SG13S153	FLA315232			A/T
29129139	29129139	SG13S97	FLA315355			A/G
29129154	29129154	SG13S154	FLA315370			C/T
29129395	29129395	SG13S40	FLA315611			G/T
29129915	29129915	SG13S155	FLA316131		rs4769875	A/G
29130192	29130192	SG13S156	FLA316408			A/C
29130256	29130256	SG13S157	FLA316472			A/G
29130299	29130299	SG13S158	FLA316515			A/C
29130353	29130353	SG13S159	FLA316569			G/T

29130391	29130391	SG13S160	FLA316607			C/T
29130547	29130547	SG13S32	FLA316763			A/C
29131280	29131280	SG13S161	FLA317496			A/G
29131403	29131403	SG13S162	FLA317619			A/G
29131404	29131404	SG13S163	FLA317620			C/T
29131431	29131431	SG13S164	FLA317647		rs4769058	C/T
29131517	29131517	SG13S165	FLA317733			A/T
29131528	29131528	SG13S166	FLA317744		rs4769059	C/T
29131599	29131599	SG13S167	FLA317815		rs4769876	A/G
29132003	29132003	SG13S168	FLA318219			A/C
29133753	29133753	SG13S33	FLA319969			G/T
29134045	29134045	SG13S41	FLA320261			A/G
29134177	29134177	SG13S169	FLA320393			A/G
29134379	29134379	SG13S404	FLA320595		rs4427651	G/T
29135558	29135558	SG13S170	FLA321774		rs3935645	C/T
29135640	29135640	SG13S171	FLA321856		rs3935644	A/G
29135750	29135750	SG13S172	FLA321966			A/G
29135809	29135809	SG13S173	FLA322025			A/T
29135877	29135877	SG13S42	FLA322093		rs4769060	A/G
29136080	29136556	EXON5				
29136290	29136290	SG13S194	FLA322506			C/T
29136462	29136462	SG13S195	FLA322678		rs1132340	A/G
29136797	29136797	SG13S174	FLA323013			A/G
29137100	29137100	SG13S34	FLA323316			G/T
29137150	29137150	SG13S175	FLA323366			A/G
29137607	29137607	SG13S176	FLA323823			A/G
29137651	29137651	SG13S177	FLA323867			C/T
29137905	29137905	SG13S178	FLA324121			C/G
29138117	29138117	SG13S35	FLA324333			A/G
29138375	29138375	SG13S179	FLA324591			A/G
29138385	29138385	SG13S180	FLA324601			C/T
29138633	29138633	SG13S181	FLA324849	DG00AAHIF	rs4420371	C/G
29139153	29139153	SG13S182	FLA325369			C/T
29139277	29139277	SG13S183	FLA325493		rs4466940	C/T
29139435	29139435	SG13S184	FLA325651	DG00AAHOI	rs4445746	A/G
29139971	29139971	SG13S185	FLA326187			A/G
29140441	29140441	SG13S405	FLA326657			A/G
29140649	29140649	SG13S91	FLA326865			A/G
29140695	29140695	SG13S186	FLA326911		rs4769877	A/T
29140703	29140703	SG13S187	FLA326919			A/G
29140805	29140805	SG13S188	FLA327021	DG00AAJFE		A/G
29141049	29141049	SG13S406	FLA327265			C/T
29142392	29142392	SG13S92	FLA328644		rs4429158	C/T
29142397	29142397	SG13S93	FLA328649			A/G
29142712	29142712	SG13S36	FLA328964			C/T
29144013	29144013	SG13S407	FLA330265			C/T
29144203	29144203	SG13S408	FLA330455			C/T
29144234	29144589	D13S1238				
29144255	29144255	SG13S7	FLA330507			C/T
29144877	29144877	SG13S37	FLA331129			A/G
29144982	29144982	SG13S409	FLA331234			A/G
29144983	29144983	SG13S8	FLA331235		rs4491352	A/C
29145122	29145122	SG13S410	FLA331374		rs4319601	C/T
29145143	29145143	SG13S411	FLA331395			A/G
29145171	29145171	SG13S9	FLA331423			C/T
29145221	29145221	SG13S412	FLA331473		rs4769062	A/G
29145265	29145265	SG13S413	FLA331517		rs4238138	C/T

minor allele	minor allele frequency (%)	start position in sequence xx	end position in sequence xx
G	10.32	432	432
G	30.46	28356	28356
T	37.38	33803	33803
G	0.545	42627	42627
G	1.111	43101	43101
G	0.328	43315	43315
C	0.495	43353	43353
A	6.993	43774	43774
C	30.876	53244	53244
G	6.731	53303	53303
T	0.353	53423	53423
C	31.356	53734	53734
A	30.935	53902	53902
G	5.492	71869	71869
A	1.812	72696	72696
G	35.00	75670	75670
C	1.314	83410	83410
T	3.521	93792	93792
A	30.031	94202	94202
A	1.724	94668	94668
A	0.369	106707	106707
A	13.66	110180	110180
A	20.779	117355	117355
T	5.965	117446	117446
A	16.923	118416	118416
A	34.364	127348	127348
A	8.537	127383	127383
T	25.536	127402	127402
		131702	131949
		132359	132753
A	37.302	134272	134272
C	6.25	138551	138551
A	0.49	149983	149983
A	14.08	150200	150200
G	0.62	150357	150357
G	14.01	151350	151350
T	0.58	151518	151518
C	30.21	153102	153102
A	10.95	153190	153190
G	30.00	154224	154224
A	27.95	155473	155473
G	2.41	156090	156090
A	0.39	156186	156186
T	10.23	156473	156473
T	15.17	157044	157044
T	13.60	157886	157886
G	12.44	158025	158025
A	13.45	158054	158054
G	14.59	158997	158997
T	26.84	159307	159307
A	12.73	159580	159580

C	43.67	159780	159780
A	12.18	160287	160287
A	8.38	160536	160536
G	0.62	160594	160594
T	12.34	160947	160947
G	25.34	161964	161964
C	0.24	162259	162259
T	25.66	162999	162999
A	14.84	164688	164688
G	12.37	164813	164813
C	14.55	164874	164874
G	11.99	164962	164962
C	14.66	165476	165476
A	12.21	165553	165553
A	0.79	166486	166486
C	10.15	166891	166891
C	3.53	166979	166979
A	12.45	169965	169965
C	0.62	171909	171909
G	31.55	172271	172271
G	4.94	172629	172629
C	15.51	172646	172646
T	27.91	173099	173099
G	14.74	174329	174329
T	1.17	174652	174652
T	1.28	175138	175138
A	2.17	175404	175404
		175668	175812
A	30.11	175830	175830
G	0.66	176398	176398
A	28.31	176579	176579
G	14.85	176919	176919
C	1.21	176972	176972
A	1.04	177112	177112
G	0.88	177182	177182
C	1.14	177344	177344
T	7.10	177557	177557
C	22.52	177773	177773
A	20.86	178096	178096
T	13.83	178178	178178
T	4.05	178508	178508
A	4.07	178630	178630
T	4.07	178689	178689
A	1.06	178862	178862
C	16.00	179889	179889
G	49.36	180174	180174
A	29.75	180264	180264
T	5.06	180306	180306
C	46.23	180455	180455
C	1.59	180583	180583
T	1.45	180680	180680
G	11.32	181139	181139
A	3.25	182056	182056
A	34.12	182738	182738
G	29.63	182940	182940
A	45.68	183878	183878

G	36.65	184020	184020
G	8.07	184068	184068
		184196	184296
T	1.02	184249	184249
A	49.57	184308	184308
A	0.58	184344	184344
C	24.71	184401	184401
T	7.19	184688	184688
A	1.10	185133	185133
T	37.65	185546	185546
A	45.50	185553	185553
T	1.22	185580	185580
T	0.89	185741	185741
T	36.69	185954	185954
T	29.11	186118	186118
A	30.19	186815	186815
G	3.29	186873	186873
T	36.96	187069	187069
G	36.63	187138	187138
T	37.34	187289	187289
C	1.15	187462	187462
T	9.91	187740	187740
C	3.36	188939	188939
T	36.24	188949	188949
A	31.58	189342	189342
G	0.45	189572	189572
T	1.14	189988	189988
T	46.57	190253	190253
A	10.34	190283	190283
T	8.00	190294	190294
T	33.71	190298	190298
T	2.29	190311	190311
G	1.19	191370	191370
A	1.01	191635	191635
A	47.88	191643	191643
		192188	192259
A	4.68	192441	192441
G	29.72	192906	192906
C	8.22	193032	193032
C	21.10	193521	193521
T	8.57	193822	193822
T	23.23	193840	193840
T	24.20	195301	195301
		196080	196162
C	23.89	196284	196284
T	19.33	196316	196316
G	11.50	196798	196798
T	3.08	197016	197016
A	9.72	197139	197139
T	0.98	197154	197154
T	2.24	197395	197395
A	1.43	197915	197915
A	1.80	198192	198192
G	2.38	198256	198256
A	0.61	198299	198299
G	2.55	198353	198353

T	0.83	198391	198391
C	48.50	198547	198547
G	2.44	199280	199280
G	2.45	199403	199403
C	2.45	199404	199404
C	2.55	199431	199431
T	20.00	199517	199517
T	2.46	199528	199528
A	3.50	199599	199599
C	8.39	200003	200003
T	8.99	201753	201753
G	5.41	202045	202045
G	4.12	202177	202177
G	38.33	202379	202379
C	32.77	203558	203558
G	48.03	203640	203640
G	1.67	203750	203750
A	0.68	203809	203809
G	42.44	203877	203877
		204080	204556
T	0.30	204290	204290
G	2.46	204462	204462
G	0.56	204797	204797
G	30.23	205100	205100
A	2.40	205150	205150
A	2.24	205607	205607
T	1.64	205651	205651
C	1.40	205905	205905
A	9.52	206117	206117
A	48.14	206375	206375
T	2.50	206385	206385
C	49.41	206633	206633
T	2.36	207153	207153
T	12.07	207277	207277
A	16.67	207435	207435
G	7.66	207971	207971
A	9.66	208441	208441
A	7.78	208649	208649
A	25.71	208695	208695
A	1.43	208703	208703
G	4.71	208805	208805
T	0.56	209049	209049
T	8.33	210392	210392
A	7.23	210397	210397
C	15.88	210712	210712
T	3.29	212013	212013
T	0.30	212203	212203
		212234	212589
T	16.28	212255	212255
G	16.70	212877	212877
A	1.93	212982	212982
C	30.64	212983	212983
T	20.57	213122	213122
A	1.54	213143	213143
C	16.37	213171	213171
A	7.42	213221	213221
T	1.91	213265	213265

Table 4

Most significant 4 microsatellite marker haplotypes in the initial haplotype analysis. Length=length of haplotype in Mb. P-val=p-value. RR=Relative risk. N af=Number of patients. P al=allelic frequency of haplotype. P ca =carrier frequency of haplotype. N ct= number of controls. Alleles= alleles in the haplotype. Markers= markers in the haplotype.

4 markers	**:**	**pos.rr-frqgt1perc**											
length	**p-val**	**RR**	**N_af**	**P_al**	**P_ca**	**N_ct**	**P_al**	**P_ca**	**Alleles**				**Markers**
0.88	4.71E-06	6.23	428	0.065	0.125	721	0.011	0.022	0	-12	-6	0	DG13S80 DG13S83 DG13S1110 DG13S163
0.82	8.60E-06	INF	438	0.032	0.062	720	0	0	0	4	2	14	DG13S111 1 DG13S1103 D13S1287 DG13S1061
0.67	6.98E-06	19.9 1	435	0.03	0.059	721	0.002	0.003	8	6	0	8	DG13S1103 DG13S163 D13S290 DG13S1061
0.767	4.85E-06	26.7 2	436	0.048	0.094	721	0.002	0.004	0	0	2	12	DG13S1101 DG13S166 D13S1287 DG13S1061
0.515	1.93E-06	INF	422	0.048	0.094	721	0	0	2	0	0	6	DG13S166 DG13S163 D13S290 DG13S1061
0.864	1.68E-06	INF	424	0.024	0.048	717	0	0	0	2	0	- 16	DG13S166 DG13S163 DG13S1061 DG13S293
0.927	5.38E-06	INF	435	0.034	0.067	720	0	0	4	2	14	3	DG13S1103 D13S1287 DG13S1061 DG13S301

Table 5

Most significant 5 microsatellite marker haplotypes in the initial haplotype analysis. Length=length of haplotype in Mb. P-val=p-value. RR=Relative risk. N af=Number of patients. P al=allelic frequency of haplotype. P ca =carrier frequency of haplotype. N ct= number of controls. Alleles= alleles in the haplotype. Markers= markers in the haplotype

5markers	:	pos.rr-frqgt1perc												
length	p-val	RR	N_af	P_al	P_ca	N_ct	P_al	P_ca	Alleles					Markers
0.851	7.45E-06	15.43	413	0.034	0.067	715	0.002	0.005	0	18	0	0	0	DG13S79 D13S1299 DG13S87 D13S1246 DG13S166
0.964	8.07E-06	INF	437	0.023	0.045	721	0	0	0	-12	6	8	6	DG13S79 DG13S83 DG13S1104 DG13S1103 DG13S163
0.964	2.38E-06	INF	437	0.026	0.052	720	0	0	0	6	0	8	6	DG13S79 DG13S1104 DG13S172 DG13S1103 DG13S163
0.931	7.05E-06	5.8	429	0.068	0.131	721	0.012	0.025	0	-6	0	0	-2	DG13S79 DG13S1110 DG13S175 DG13S166 D13S1238
0.964	8.13E-06	INF	434	0.021	0.041	721	0	0	0	3	8	2	6	DG13S79 DG13S1098 DG13S1103 DG13S166 DG13S163
0.597	9.78E-06	4.58	428	0.074	0.143	717	0.017	0.034	-6	0	0	0	-2	DG13S1110 DG13S89 DG13S175 DG13S166 D13S1238
0.896	6.92E-06	INF	428	0.026	0.051	721	0	0	-12	-6	0	-2	2	DG13S83 DG13S1110 DG13S166 D13S1238 D13S290
0.722	2.18E-06	INF	453	0.026	0.051	738	0	0	-6	0	0	-2	2	DG13S1110 D13S289 DG13S166 D13S1238 D13S290
0.982	7.88E-06	INF	437	0.028	0.055	721	0	0	0	0	4	2	14	DG13S87 DG13S175 DG13S1103 D13S1287 DG13S1061

0.841	8.88E-06	INF	438	0.032	0.062	720	0	0	0	0	4	2	14	DG13S89 DG13S1111 DG13S1103 D13S1287 DG13S1061
0.841	9.67E-07	INF	435	0.029	0.057	721	0	0	0	8	6	0	8	DG13S89 DG13S1103 DG13S163 D13S290 DG13S1061
0.982	7.90E-06	18.63	437	0.026	0.052	721	0.001	0.003	0	4	0	2	14	DG13S87 DG13S1103 DG13S166 D13S1287 DG13S1061
0.841	3.52E-06	28.52	436	0.048	0.094	721	0.002	0.004	0	0	0	2	12	DG13S89 DG13S1101 DG13S166 D13S1287 DG13S1061
0.705	5.28E-06	INF	435	0.027	0.053	721	0	0	0	8	6	0	8	DG13S175 DG13S1103 DG13S163 D13S290 DG13S1061
0.841	4.21E-06	INF	422	0.048	0.093	721	0	0	0	2	0	0	6	DG13S89 DG13S166 DG13S163 D13S290 DG13S1061
0.767	4.02E-06	28.11	436	0.049	0.095	721	0.002	0.004	0	0	0	2	12	DG13S1101 DG13S175 DG13S166 D13S1287 DG13S1061
0.767	1.29E-06	31.07	436	0.047	0.092	721	0.002	0.003	0	0	0	2	12	DG13S1101 DG13S172 DG13S166 D13S1287 DG13S1061
0.705	4.25E-07	INF	422	0.048	0.093	721	0	0	0	2	0	0	6	DG13S175 DG13S166 DG13S163 D13S290 DG13S1061
0.683	6.58E-06	INF	437	0.029	0.056	721	0	0	0	4	0	2	14	DG13S172 DG13S1103 DG13S166 D13S1287 DG13S1061
0.767	2.85E-06	32.43	436	0.044	0.087	721	0.001	0.003	0	0	0	2	12	DG13S1101 DG13S166 D13S290 D13S1287 DG13S1061
0.865	9.58E-06	18.39	451	0.023	0.045	739	0.001	0.003	0	0	2	2	-16	D13S289 DG13S166 DG13S163 D13S1287 DG13S293
0.865	5.08E-06	INF	453	0.019	0.038	739	0	0	0	0	2	0	-16	D13S289 DG13S166 DG13S163 DG13S1061 DG13S293
0.927	1.02E-07	27.65	437	0.037	0.073	721	0.001	0.003	4	0	2	14	3	DG13S1103 DG13S166

														D13S1287 DG13S1061 DG13S301

Table 6

Position (Mb) of microsatellite markers sequence assembly (SA5), primers and size of the markers.

mb	marker	Forward	Reverse	size
25.092042	DG13S2101	ACGGTGATGACGCCTACATT (SEQ ID NO: 4)	TCACATGGACCAATTACCTAGAA(SEQ ID NO: 5)	188
25.092042	DG13S48	CAAATTTCAGATGTGCCAACC (SEQ ID NO: 6)	ACGGTGATGACGCCTACATT(SEQ ID NO: 7)	214
25.396504	D13S1304	ACCAGCCTTTGCTTAGGA(SEQ ID NO: 8)	ACATTCTAGTGCTACAGGGTACTC(SEQ ID NO: 9)	133
25.396535	DG13S2105	TGTTCTGCACACGAACATTCT(SEQ ID NO: 10)	TCCTGAGTCCTCTCCACCTG(SEQ ID NO: 11)	104
25.445511	DG13S2106	TGGGAATTAATGAAGAACAACAAA(SEQ ID NO: 12)	CATGTTTCGAAGAACTCAAGAGG(SEQ ID NO: 13)	428
25.544920	D13S1254	AAATTACTTCATCTTGACGATAACA(SEQ ID NO: 14)	CTATTGGGGACTGCAGAGAG (SEQ ID NO: 15)	218
25.544925	DG13S2107	GGGACTGCAGAGAGCAGAAG (SEQ ID NO: 16)	CAAGAAGGGAAATTCCTACGC (SEQ ID NO: 17)	95
25.565956	DG13S55	AGCCAGTGTCCACAAGGAAG (SEQ ID NO: 18)	GAGGGTGAGACACATCTCTGG (SEQ ID NO: 19)	283
25.605793	DG13S54	AATCGTGCCTCAGTTCCATC (SEQ ID NO: 20)	CCACCAGGAACAACACACAC (SEQ ID NO: 21)	156
25.619693	D13S625	TTGCTCTCCAGCCTGGGC (SEQ ID NO: 22)	TTCCTCTGGCTGCCTGCG (SEQ ID NO: 23)	185
25.687422	DG13S1479	TTTGATTCCGTGGTCCATTA (SEQ ID NO: 24)	TTATTTGGTCGGTGCACCTTT (SEQ ID NO: 25)	339
25.749344	DG13S1440	GGTAGGTTGAAATGGGCTAACA (SEQ ID NO: 26)	TCATGACAAGGTGTTGGATTT (SEQ ID NO: 27)	153
25.9	DG13S18	CCTCCTCTGCCATGAAGCT	CTATTTGGTCTGCGGGTT	418

01212	90	A (SEQ ID NO: 28)	GT (SEQ ID NO: 29)	
25.92808 1	DG13S18 79	TTTGAGCCCAGATCTAAGC AA (SEQ ID NO: 30)	AAATGTTAATGTCACCGA CAAA (SEQ ID NO: 31)	443
25.93260 9	DG13S15 40	TACTGGGTTATCGCCTGAC C (SEQ ID NO: 32)	CCAATGGACCTCTTGGAC AT (SEQ ID NO: 33)	152
25.94674 3	DG13S18 89	TTTGAATGTTCATATATTTG TGGTG (SEQ ID NO: 34)	CCCTCGTAATGAAACCTA TTTGA (SEQ ID NO: 35)	222
25.94867 9	DG13S59	TTTCGGCACAGTCCTCAAT A (SEQ ID NO: 36)	CAGGGTGTGGTGACAT (SEQ ID NO: 37)	228
25.95234 7	DG13S18 94	TGTTTCTTTCTTTCTCTCTC TCTTTC (SEQ ID NO: 38)	AAATGAGTTCAATGAGTT GTGGTT (SEQ ID NO: 39)	209
25.98836 0	DG13S15 45	CAGAGAGGAACAGGCAGA GG (SEQ ID NO: 40)	AGTGGCTGGGAAGCCTT ATT (SEQ ID NO: 41)	394
26.07186 6	DG13S15 24	AGGTGAGAGAACAAACCTG TCTT (SEQ ID NO: 42)	GCCTTCCTTCTAAGGCCA AC (SEQ ID NO: 43)	115
26.18349 2	DG13S14 91	TGTTATACATTTCAATTTCA CCTCA (SEQ ID NO: 44)	GTACTCCAGCCGGGCAA C (SEQ ID NO: 45)	286
26.23628 9	DG13S62	TTGTTCAGTGCTCTATAGTT ACAAAGT (SEQ ID NO: 46)	GGTCACAAAGCTATGCGA TTA (SEQ ID NO: 47)	158
26.27346 3	D13S1244	TCAACAAGTGGATTAAGAA ACTGTG (SEQ ID NO: 48)	CTGTTTATGGCTGAGAAG TATGC (SEQ ID NO: 49)	86
26.28693 5	DG13S64	TAGCAGGGTGCAGTCTA (SEQ ID NO: 50)	ACCATACCACCACCACCA TC (SEQ ID NO: 51)	247
26.31450 1	D13S243	ACTGTACTTCTGCCTGGGC (SEQ ID NO: 52)	TTTTGTAATGCCTCAACC ATG (SEQ ID NO: 53)	147
26.32718 4	DG13S15 29	CTGTAGACTTTATCCCTGA CTTACTG (SEQ ID NO: 54)	CAATGAATGATGAAGATT CCACTC (SEQ ID NO: 55)	132
26.33876 7	DG13S19 08	TGACACCATGTCTTACTGT TTGC (SEQ ID NO: 56)	GAGGATACAATGAGAACC AAATCTC (SEQ ID NO: 57)	224

26.38803 4	DG13S15 46	CCACAGAATGCTCCAAAGG T (SEQ ID NO: 58)	GAGTTCAAGTGATGGATG ACGA (SEQ ID NO: 59)	357
26.43581 1	DG13S14 44	CAGATAGATGAATAGGTGG ATGGA (SEQ ID NO: 60)	CACTGTTCCAAGTGCTTT GC (SEQ ID NO: 61)	193
26.48665 7	DG13S14 58	GCAGGGCAAACTGCCTTAT (SEQ ID NO: 62)	TTTGGTGAAATGTCTGTT TATGG (SEQ ID NO: 63)	402
26.50454 5	D13S252	CTCAACCTGGCTTCTACT (SEQ ID NO: 64)	TACTCCTTAATAAACTCC CC (SEQ ID NO: 65)	338
26.50823 1	DG13S66	TATGCGTTGTGTGTGTG (SEQ ID NO: 66)	GGGCCTTAGATTCTTGTA GTGG (SEQ ID NO: 67)	217
27.11512 0	DG13S15 54	CTCGCATCTCGCTTCTCAC T (SEQ ID NO: 68)	CTCAAGGGTCCAGTGGTT TG (SEQ ID NO: 69)	420
27.14067 5	DG13S19 07	TGTCCAGACTGCCTCCTAC A (SEQ ID NO: 70)	TGCAACACCTGGTTCACA AT (SEQ ID NO: 71)	131
27.14584 2	D13S802	CACAGTGAGACTCTATCTC AAAAA (SEQ ID NO: 72)	TCAGACTGGCTTAGACTG TGG (SEQ ID NO: 73)	150
27.24061 6	DG13S18 92	AAATTCCAAAGGCCAGGTG (SEQ ID NO: 74)	CCATACAGTTTCCTAGGT TCTGG (SEQ ID NO: 75)	373
27.25345 2	DG13S18 49	CACCTGGCCAAATGTTTGT T (SEQ ID NO: 76)	TGCTTGAATCCAGAGACT GC (SEQ ID NO: 77)	190
27.27386 0	DG13S68	TTTGCGAGTCCTTGTGGAG T (SEQ ID NO: 78)	ACAGTCCGCTCCCTCCTA AT (SEQ ID NO: 79)	238
27.28046 1	DG13S69	ATGCTTGGCCCTCAGTTT (SEQ ID NO: 80)	TTGGCAACCCAAGCTAAT ATG (SEQ ID NO:81)	296
27.48379 9	D13S1250	CTCCACAGTGACAGTGAG G (SEQ ID NO:82)	GAGAGGTTCCCAATCCC (SEQ ID NO: 83)	160
27.61040 6	D13S1448	CATCAACCTCCCCACCAC (SEQ ID NO: 84)	TATTTTTTCAGTCCCACA GTTAGC (SEQ ID NO: 85)	227
27.61581	DG13S57 4	CAGCTCCTGGCCATATTTC T (SEQ ID NO: 86)	GAGCCATTTCTCTGGGTC TG (SEQ ID NO:87)	153

4				
27.641211	DG13S73	GGTCCGTGTCAACCCTTAGA (SEQ ID NO: 88)	CAGGTTGATGGGAGGGAAA (SEQ ID NO: 89)	198
27.661507	DG13S1532	CGGGAAATGACAGTGAGACC (SEQ ID NO: 90)	TGCCTAGATTCTCCCGTAAG (SEQ ID NO: 91)	163
27.705347	D13S1242	GTGCCCAGCCAGATTC (SEQ ID NO: 92)	GCCCCCAGTCAGGTTT (SEQ ID NO: 93)	198
27.883872	DG13S576	TTTCTCTCTCCACGGAATGAA (SEQ ID NO: 94)	AACCCATTCTCACAGGGTGTA (SEQ ID NO: 95)	199
27.897365	DG13S1917	AGGAGTGTGGCAGCTTTGAG (SEQ ID NO: 96)	TGGATTCCCGTGAGTACCAG (SEQ ID NO: 97)	165
27.932154	D13S217	ATGCTGGGATCACAGGC (SEQ ID NO: 98)	AACCTGGTGGACTTTTGCT (SEQ ID NO: 99)	170
28.080632	DG13S581	AGCATTTCCAATGGTGCTTT (SEQ ID NO: 100)	CATGTTGATATGCCTGAAGGA (SEQ ID NO:101)	367
28.165348	DG13S1471	CACTGTCTGCTGCCACTCAT (SEQ ID NO:102)	AGAGATTATGTGATGTACCCTCTCTAT(SEQ ID NO:103)	267
28.303252	DG13S583	CAAGCCTGGGACACAGAAAT (SEQ ID NO: 104)	TTTGCAGACACCACAACACA (SEQ ID NO: 105)	264
28.303256	D13S120	ATGACCTAGAAATGATACTGGC (SEQ ID NO: 106)	CAGACACCACAACACACATT (SEQ ID NO: 107)	175
28.385566	D13S1486	TGGTTTAAAAACCTCATGCC (SEQ ID NO: 108)	ATCCCAAACTCTGTACTTATGTAGG (SEQ ID NO: 109)	151
28.415530	DG13S1024	TTTGCACATACACATAAGCGAAC (SEQ ID NO: 110)	CACAAATCCCGTGCACTAAA (SEQ ID NO: 111)	139
28.415530	DG13S1510	ATTCCTGGGCTCATGGTACA (SEQ ID NO: 112)	TGCCGTCATCTGCTTTAGAA (SEQ ID NO: 113)	390
28.430308	DG13S1495	CCTTGGCTGTTGTGACTGGT (SEQ ID NO: 114)	CACTCAGGTGGGAGGATCAC (SEQ ID NO: 115)	285
28.5	DG13S14	GCTGTTTCCTTGGCTTCTT	CCCATACTTGAGATGACC	291

1754 1	82	CT (SEQ ID NO: 116)	ATGA (SEQ ID NO: 117)	
28.5 5106 0	DG13S18 45	CACTTTGCCAGTAGCCTTG A (SEQ ID NO: 118)	TTGGGAAAGTTAACCCAG AGA (SEQ ID NO: 119)	284
28.6 3490 3	DG13S10 30	TTTGGGAAGAGCCATGAGA C (SEQ ID NO: 120)	CTCTGGGCATTGGAGGA TTA (SEQ ID NO: 121)	354
28.6 3490 3	DG13S14 67	TTTGGGAAGAGCCATGAGA C (SEQ ID NO: 122)	AATGCCCATGTGCACTGT AG (SEQ ID NO: 123)	231
28.6 8660 7	DG13S58 4	GGGAGACAAGTCAGGTGA GG (SEQ ID NO: 124)	CTGAGTATGGAGTCTTCA TCATTATC (SEQ ID NO: 125)	151
28.7 9403 2	DG13S15 19	TCGTCTCGAAGAAAGAAAG AAGA (SEQ ID NO:126)	CACCATGGTTAATTGCA CA (SEQ ID NO: 127)	286
28.8 7615 6	DG13S77	TGACGTGGGTTCAGGTTGT A (SEQ ID NO: 128)	AGTGCATTGGTGCCTTCT CT (SEQ ID NO: 129)	220
28.9 7072 3	DG13S58 6	GGACTGCCAATTCTACAGC A (SEQ ID NO: 130)	TTTCCATGGGAAATTTGG TC (SEQ ID NO: 131)	151
28.9 7564 1	DG13S79	TGCTACTAGATTTGACCAA CCA (SEQ ID NO: 132)	GACTTGTAAAGGATTTAG TGATTTCG (SEQ ID NO: 133)	128
29.0 5939 4	DG13S80	GTGGAAGGCCTCTCTTG (SEQ ID NO: 134)	TGCTTCTTGAGGGAAAGC AT (SEQ ID NO: 135)	233
29.1 2615 2	DG13S82	CACGTGGTTCACCTCTCTA GG (SEQ ID NO: 136)	TTGGCCACTTATTTGTG (SEQ ID NO: 137)	302
29.1 5469 1	D13S1299	CGATGAGTGACAGGGCT (SEQ ID NO: 138)	CCTCGTGGGTGGAATAA (SEQ ID NO: 139)	225
29.1 5473 7	DG13S85	TTGGCCATTAGCAATTAGC A (SEQ ID NO: 140)	CGTGGGTGGAATAAATCA GG (SEQ ID NO: 141)	153
29.1 5846 2	D13S629	GTTGAGGCAAGAGAATCAC T (SEQ ID NO: 142)	GCACATTTACACCAGGGT G (SEQ ID NO.143)	145
29.2 2406 0	DG13S19 34	CCTTCAGAGGATTTCCCTT TC (SEQ ID NO: 144)	CTGGTTTGACTCCAGCTT CA (SEQ ID NO: 145)	431

29.245462	DG13S1098	TGTTCAAACCTAAGGTGCTTCA (SEQ ID NO: 146)	GAAACAACAACAACAACAACAACA (SEQ ID NO: 147)	416
29.259840	DG13S1104	CCTGGCACGGAATAGACACT (SEQ ID NO: 148)	GGCCTCCTTTGCTCTGAAG (SEQ ID NO: 149)	378
29.294436	DG13S1097	CATCCCTGTGGCTGATTAAGA (SEQ ID NO: 150)	AACAGTTCCAGCCCGTTCTA (SEQ ID NO: 151)	162
29.309700	DG13S1110	TTTCAAAGGAATATCCAAGTGC (SEQ ID NO: 152)	TGGCGTACCATATAAACAGTTCTC (SEQ ID NO: 153)	265
29.309909	DG13S86	TTTCAAAGGAATATCCAAGTGC (SEQ ID NO: 154)	AAACGTGACACTTCCACACA (SEQ ID NO: 155)	177
29.359961	DG13S87	TTCAATGAAGGTGCCGAAGT (SEQ ID NO: 156)	TGTCTATCCCAAAGCAA (SEQ ID NO: 157)	218
29.522443	DG13S1111	GCAAGACTCTGTTGAAGAAGAAGA (SEQ ID NO: 158)	TCCCTCTGTTTGAGTTTCTCG (SEQ ID NO: 159)	110
29.574665	DG13S1101	AGGCACAGTCGCTCATGTC (SEQ ID NO: 160)	AAACTTTAGCTAATGGTGGTCAAA (SEQ ID NO. 161)	333
29.622755	DG13S1106	TGTGATTCCAGGGAGCTATCA (SEQ ID NO. 162)	TAGGTGTGTGGAGGACAGCA (SEQ ID NO. 163)	416
29.658910	DG13S172	CCAGTTTCAGTTAGCCAAGTCTG (SEQ ID NO: 164)	GAGAGGGAATGAATGCAGGA (SEQ ID NO: 165)	267
29.665709	D13S1246	GAGCATGTGTGACTTTCATATTCAG(SEQ ID NO: 166)	AGTGGCTATTCATTGCTACAGG(SEQ ID NO: 167)	177
29.672561	*DG13S1103*	TTGCTGGATGCTGGTTTCTA(SEQ ID NO: 168)	AAAGAGAGAGAGAAAGAGAAAGAAAGA(SEQ ID NO: 169)	264
29.825975	D13S289	CTGGTTGAGCGGCATT(SEQ ID NO: 170)	TGCAGCCTGGATGACA(SEQ ID NO: 171)	260
29.826631	*DG13S166*	CCTATGGAAGCATAGGGAAGAA(SEQ ID NO: 172)	CCCACTTCTGAGTCTCCTGAT(SEQ ID NO: 173)	395
29.90668	DG13S164	GGGATGCAGAAAGGATGTGT(SEQ ID NO: 174)	AAGAATGCTGGCCAACGTAA(SEQ ID NO: 177)	218

9				
29.90670 0	D13S1238	CTCTCAGCAGGCATCCA(SEQ ID NO: 178)	GCCAACGTAATTGACACCA(SEQ ID NO:179)	129
30.03137 8	D13S290	CCTTAGGCCCCATAATCT(SEQ ID NO: 180)	CAAATTCCTCAATTGCAAAAT(SEQ ID NO:181)	176
30.08630 3	D13S1229	GGTCATTCAGGGAGCCATTC(SEQ ID NO: 182)	CCATTATATTTCACCAAGAGGCTGC(SEQ ID NO: 183)	119
30.19284 7	DG13S1460	TGCCTGGTCATCTACCCATT(SEQ ID NO: 184)	TCTACTGCAGCGCTGATCTT(SEQ ID NO: 185)	264
30.21767 0	DG13S1933	CATTTATGAATGGAGGTGAAGC(SEQ ID NO: 186)	ATGGGAGCTCAAAGGGAAAT(SEQ ID NO: 187)	186
30.30321 3	DG13S1448	CAGCAGGAAGATGGACAGGT(SEQ ID NO: 188)	CACACTGCATCACACATACCC(SEQ ID NO: 189)	136
30.31787 1	*D13S1287*	TATGCCAGTATGCCTGCT(SEQ ID NO: 190)	GTCACATCAGTCCATTTGC(SEQ ID NO: 191)	232
30.34210 2	*DG13S1061*	CCAAAGCAAGTAACCTCCTCA(SEQ ID NO: 192)	AAACAATCACTGCCCTCTGG(SEQ ID NO. 193)	227
30.57183 7	DG13S1904	TGATGAAATTGCCTAGTGATGC(SEQ ID NO: 194)	GGATCCAATCGTACGCTACC(SEQ ID NO. 195)	136
30.64343 8	DG13S882	CGAATGGGTGACTAACAGCA(SEQ ID NO: 196)	CTGGAGTGCAGGGACATGA(SEQ ID NO: 197)	378
30.66593 7	DG13S295	AAAGAAATATTCCAAGAAGAAAGAAA(SEQ ID NO: 198)	TTGCACAACTTTGTGTAGAGCAT(SEQ ID NO: 199)	279
30.67446 8	D13S1226	GGGTATGTCTTTATTCTCGGCAGTA(SEQ ID NO: 200)	GTGCATTCACAGACCAGTCATT(SEQ ID NO: 201)	219
30.69095 9	DG13S293	GGGCTTGAAGGCACTAAATGT(SEQ ID NO: 202)	CCAAGCAGTAATTCCTTCCTCA(SEQ ID NO:203)	313
30.71246 8	DG13S1490	ACCTAAACACCACGGACTGG(SEQ ID NO: 204)	CAGGTATCGACATTCTTCCAAA(SEQ ID NO: 205)	418
30.8	DG13S93	TGGGAAGCCAGTAAAGTAG	AAAGAGACTCCACACATC	190

24483		GAA(SEQ ID NO: 206)	CATTT(SEQ ID NO: 207)	
30.824859	DG13S94	AGGGCTATTCCTCAAGGTGTT(SEQ ID NO: 208)	TGCTAACACTACCCTCGCAAT(SEQ ID NO: 209)	332
30.928429	DG13S1534	GGGCAGGAATCTCTGAAGTG (SEQ ID NO: 210)	CTCCACTGAGAAGCCAAGGA(SEQ ID NO. 211)	382
30.940369	DG13S95	AGGCCAAGCTGGTCCATAG(SEQ ID NO: 212)	TCTCTCAAAGCCTCGCTCTC(SEQ ID NO: 213)	126
30.970238	DG13S96	CCTTTGAGGCTGGATCTGTT(SEQ ID NO: 214)	TTTCCTTATCATTCATTCCCTCA(SEQ ID NO: 215)	218
31.038874	D13S260	AGATATTGTCTCCGTTCCATGA(SEQ ID NO: 216)	CCCAGATATAAGGACCTGGCTA(SEQ ID NO: 217)	163
31.092294	DG13S17	TTTAAGCCCTGTGGAATGTATTT(SEQ ID NO: 218)	GACATTGCAGGTCAAGTAGGG(SEQ ID NO: 219)	157
31.207844	DG13S306	TGCATAAGGCTGGAGACAGA(SEQ ID NO: 220)	CACAGCAGATGGGAGCAAA(SEQ ID NO:221)	158
31.260521	DG13S18	GTGCATGTGCATACCAGACC(SEQ ID NO: 222)	GGCAAGATGACCTCTGGAAA(SEQ ID NO: 223)	319
31.299720	DG13S1905	GTCCACTGCAGCACACAGAG(SEQ ID NO: 224)	GCACTGGTAGATACATGCTAACG(SEQ ID NO: 225)	383
31.353230	DG13S307	GGGTATCTTGGCCAGGTGT(SEQ ID NO: 226)	TGGCTAAGCACAATCCCTTT(SEQ ID NO: 227)	403
31.355135	DG13S1062	TTTGTGTTCCAGGTGAGAATTG(SEQ ID NO: 228)	GAACCATATCCCAAGGCACT(SEQ ID NO: 229)	120
31.414329	DG13S1874	AACCCAAATCAACAAACCAGA(SEQ ID NO: 230)	AATGAATTCTGGGTCACATGC(SEQ ID NO: 231)	404
31.429562	DG13S1093	TTGTTCCCACATTCATTCTACA(SEQ ID NO: 232)	TTAAACTCGTGGCAAAGACG(SEQ ID NO: 233)	273
31.626502	DG13S1059	CACCATGCCTGGCTCTTT(SEQ ID NO: 234)	AACTTCTCCAGTTGTGTGGTTG (SEQ ID NO: 235)	330

31.72374 9	DG13S10 86	AGCTGAGCTCATGCCACT(SEQ ID NO: 236)	CAAGACCTTGTGCATTTG GA(SEQ ID NO: 237)	155
31.74607 4	DG13S15 15	AGCCAGACATGGTAGTGTG C(SEQ ID NO: 238)	GCAATAACTCACACATCA GCAA (SEQ ID NO:239)	417
31.85573 2	D13S171	CCTACCATTGACACTCTCA G(SEQ ID NO: 240)	TAGGGCCATCCATTCT(S EQ ID NO: 241)	231
31.91733 2	DG13S10 92	ACCAAGATATGAAGGCCAA A(SEQ ID NO: 242)	CCTCCAGCTAGAACAATG TGAA(SEQ ID NO: 243)	176
32.00285 2	DG13S14 49	TGTCCATAGCTGTAGCCCT GT(SEQ ID NO: 244)	CTCAATGGGCATCTTTAG GC(SEQ ID NO: 245)	279
32.07295 7	DG13S14 89	TGTAATTCAACGACTGGTG TCC(SEQ ID NO: 246)	AGCTTCTGATGGTTGCTG GT(SEQ ID NO: 247)	130
32.08398 9	DG13S31 2	CAAACAAACAAACAAGCAA ACC(SEQ ID NO: 248)	TGGACGTTTCTTTCAGTG AGG(SEQ ID NO: 249)	349
32.12517 7	DG13S15 11	TGATAACTTACCAGCATGT GAGC(SEQ ID NO: 250)	TCACCTCACCTAAGGATC TGC(SEQ ID NO: 251)	314
32.18354 7	DG13S31 4	CATGCAATTGCCCAATAGA G(SEQ ID NO: 252)	TTGGGCTTGTCTACCTAG TTCA(SEQ ID NO: 253)	335
32.19535 8	DG13S10 90	TGGGTTCCTCATACTGGAG TG(SEQ ID NO: 254)	GCCTGAGCTCCAAGCTCT TT(SEQ ID NO: 255)	169
32.25103 8	DG13S10 71	GCTGCACGTATTTGTTGGT G(SEQ ID NO: 256)	AAACAGCAGAAATGGGAA CC(SEQ ID NO: 257)	239
32.35689 5	DG13S10 68	CCGTGGGCTATCAATTTCT G(SEQ ID NO: 258)	AAGATGCAATCTGGTTTC CAA(SEQ ID NO: 259)	238
32.37304 0	DG13S10 77	CCCAAGACTGAGGAGGTC AA(SEQ ID NO: 260)	GCTGACGGAGAGGAAAG AGA(SEQ ID NO: 261)	374
32.42278 0	DG13S19 06	TGACAAGGGTGTGGTTATG G (SEQ ID NO: 262)	CCGCACTTTCTCTTCTGG AC (SEQ ID NO:263)	425
32.51159	DG13S31 6	TGAGAAGCCTGGGCATTAA G (SEQ ID NO: 264)	ACAAGCTCATCCAGGGAA AG (SEQ ID NO: 265)	243

0				
32.61051 7	DG13S317	TTGGAAAGGAAGAAAGGAAGG (SEQ ID NO: 266)	TTGAAACCTAAATGCCACCTG (SEQ ID NO:267)	215
32.61071 3	D13S1493	ACCTGTTGTATGGCAGCAGT (SEQ ID NO: 268)	GGTTGACTCTTTCCCCAACT (SEQ ID NO: 269)	248
32.78989 4	DG13S1558	AGAGCTGATCTGGCCGAAG (SEQ ID NO: 270)	GGTGGACACAGAATCCACACT (SEQ ID NO: 271)	399
32.86595 0	D13S267	GGCCTGAAAGGTATCCTC (SEQ ID NO: 272)	TCCCACCATAAGCACAAG (SEQ ID NO: 273)	160
32.96141 0	DG13S1478	TCAACCTAGGATTGGCATTACA (SEQ ID NO: 274)	TCTAGGATTTGTGCCTTTCCA (SEQ ID NO: 275)	387
33.00992 2	DG13S1513	GACGTCTTAGGATTGACTTCTGC (SEQ ID NO: 276)	CCAAATACACATTCTTAAAGGGAAA (SEQ ID NO: 277)	173
33.12569 6	DG13S1461	GACTGCAGATCGTGGGACTT (SEQ ID NO: 278)	TTCTCCAGAGAAACCAAACCA (SEQ ID NO: 279)	148
33.16846 8	DG13S1551	ATTCGTGCAGCTGTTTCTGC (SEQ ID NO: 280)	GCATGACATTGTAAATGGAGGA (SEQ ID NO:281)	263
33.25498 9	DG13S1884	GGTGGGAATGTGTGACTGAA (SEQ ID NO: 282)	CCAGGTACAACATTCTCCTGAT (SEQ ID NO:283)	123
33.34012 4	D13S1293	TGCAGGTGGGAGTCAA (SEQ ID NO: 284)	AAATAACAAGAAGTGACCTTCCTA (SEQ ID NO: 285)	129
33.34690 8	DG13S326	TGTTCTCCTCACCCTGCTCT (SEQ ID NO: 286)	TTTCAGGCTAGGAAGATCCTTT (SEQ ID NO: 287)	261
33.39262 9	DG13S1518	AAAGGATGCATTCGGTTAGAG (SEQ ID NO: 288)	ACTGTCCTGTGCCTGTGCTT (SEQ ID NO: 289)	375
33.40552 7	DG13S23	CCTGAATAGGTGGAATTAAGATCAA (SEQ ID NO: 290)	TCAAGGAGCATACACACACACA (SEQ ID NO: 291)	107
33.43153 6	D13S620	GTCCACCTAATGGCTCATTC (SEQ ID NO: 292)	CAAGAAGCACTCATGTTTGTG (SEQ ID NO: 293)	185
33.4	DG13S18	AGCCTGTGATTGGCTGAGA	GGCTTACAGCTGCCTCCT	410

3709 2	66	(SEQ ID NO: 294)	TT (SEQ ID NO: 295)	
33.4 9571 8	DG13S19 27	CCCACAGAGCACTTTGTTA GA (SEQ ID NO: 296)	GCCTCCCTTAAGCTGTTA TGC (SEQ ID NO: 297)	401
33.5 0344 0	DG13S15 03	CACTCTTTACTGCCAATCA CTCC (SEQ ID NO:298)	GCCGTGTGGGTGTATGA AT (SEQ ID NO: 299)	226
33.5 6810 0	DG13S33 2	TTGTACCAGGAACCAAAGA CAA (SEQ ID NO: 300)	CACAGACAGAGGCACATT GA (SEQ ID NO: 301)	176
33.6 7584 1	DG13S33 3	GCTCTGGTCACTCCTGCTG T (SEQ ID NO: 302)	CATGCCTGGCTGATTGTT T (SEQ ID NO: 303)	446
33.7 7138 9	D13S220	CCAACATCGGGAACTG (SEQ ID NO: 304)	TGCATTCTTTAAGTCCAT GTC (SEQ ID NO: 305)	191
33.8 1804 1	DG13S19 19	CAGCAACTGACAACTCATC CA (SEQ ID NO: 306)	CCTCAATCCTCAGCTCCA AC (SEQ ID NO.307)	255
33.8 7361 4	DG13S14 39	TCCTTCACAGCTTCAAACT CA (SEQ ID NO: 308)	AGTGAGAAGCTTCCATAC TGGT (SEQ ID NO: 309)	239
33.9 0606 5	DG13S33 5	GCCAACCGTTAGACAAATG A (SEQ ID NO: 310)	CTACATGTGCACCACAAC ACC (SEQ ID NO: 311)	201
33.9 2865 3	DG13S34 0	AGTTTATTGCCGCCGAGAG (SEQ ID NO. 312)	ACCCACCACATTCACAAG C (SEQ ID NO: 313)	373
34.0 1945 5	DG13S14 96	CGATTGCCATGTCTCTTTG A (SEQ ID NO: 314)	GAGATCTGGCCTGGATTT GT (SEQ ID NO: 315)	155
34.0 3408 9	DG13S34 2	TGAGGCCAGCCTTACCTCT AT (SEQ ID NO: 316)	CCAGACATGGTGGCTTGT (SEQ ID NO: 317)	366
34.0 6177 7	DG13S34 4	GAAGGAAGGAAGGGAAGG AA (SEQ ID NO: 318)	AAGGATGAGAAGAGTCC ATGC (SEQ ID NO: 319)	292
34.0 6723 9	DG13S34 5	AAATACCCTTTGAACAGAC ACAC (SEQ ID NO: 320)	TAGCTGAGCATGGTGGTA CG (SEQ ID NO: 321)	201
34.0 7787 4	DG13S34 6	AAAGACAAGACAGCAATCC AAA (SEQ ID NO: 322)	GCAGAACCCAGGCTACA GAT (SEQ ID NO. 323)	152

34.08413 8	DG13S34 7	TCATTGTCAGCACAGAATG AACT(SEQ ID NO: 324)	GGAGGGAGGGAAGAAAG AGA (SEQ ID NO: 325)	338
34.08432 6	D13S624	GCAACACAGTGAAAGCCCA (SEQ ID NO: 326)	ACAGGAGCATGCCACCA TG(SEQ ID NO: 327)	191
34.15607 5	DG13S33 9	GGGAAGAGGAGATTGACTT GTT(SEQ ID NO: 328)	GGAACACCATCATTCCAA CC(SEQ ID NO: 329)	232
34.19247 8	DG13S19 26	TACAAGCTCCACCGTCCTT C(SEQ ID NO: 330)	TGAGTTGCTGCCTCTTCA AA(SEQ ID NO: 331)	261
34.22022 7	DG13S14 69	TGCTAATGGGCCAAGGAAT A(SEQ ID NO: 332)	GCTAAATGTCCTCATGAA TAGCC(SEQ ID NO: 333)	382
34.30144 8	DG13S35 1	TGTCCTGCAGACAGATGGT C(SEQ ID NO: 334)	CCTCCGGAGTAGCTGGA TTA(SEQ ID NO: 335)	294
34.38788 3	DG13S26	GAGACTGGCCCTCATTCTT G(SEQ ID NO: 336)	AAGAAGCCAGAGACAAA GAAATACA(SEQ ID NO: 337)	330
34.53544 1	DG13S30	CATCTATCTTTGGATTCAGT GGTG(SEQ ID NO: 338)	TGCTCCCAACATCTTACC AG(SEQ ID NO: 339)	388
34.56559 4	DG13S14 35	TGTCCTCTGGTCATTTCTAT GGT(SEQ ID NO: 340)	CATGAATGAGAAGTGATG AATGG (SEQ ID NO: 341)	235
34.65985 8	DG13S14 46	AACACGGGAAATTCCAACA G(SEQ ID NO: 342)	TGAAGAACTGAAATTGCC AGTAA(SEQ ID NO: 343)	379
34.71226 0	DG13S35 6	CAGACACTGTAAACTGGCT TCG(SEQ ID NO: 344)	GCCACATTGCTATCAGCG TA(SEQ ID NO: 345)	212
34.73875 6	DG13S35 7	TGTCATAGGCTTGCGGTAT TT(SEQ ID NO: 346)	TTGGTAGGGTCCTTTCCT TT(SEQ ID NO: 347)	202
34.77057 1	DG13S10 32	GCCTGCTCACTGTTGTTTG A(SEQ ID NO: 348)	CGGTTATCAGAGACTGGT GGT(SEQ ID NO: 349)	211
34.79967 9	DG13S15 57	GGCTTATTTCATGTACGGC TA(SEQ ID NO: 350)	GGTTAAACTCTACTTAGT CCTGATGC(SEQ ID NO: 351)	158
34.88293	DG13S19 25	GAACTCTGCAGGCACCTCT T(SEQ ID NO: 352)	CCTGAAGCGCTTGTACTG AA(SEQ ID NO: 353)	456

4				
34.9 3269 0	DG13S14 84	TGTTGCGTACTCAGCCCAT A (SEQ ID NO:354)	GACAGGTGTCAAACGGG TCT(SEQ ID NO: 355)	246
34.9 4254 7	DG13S36 0	TTGGCTTCTCGCTCTTTCTT (SEQ ID NO: 356)	AGCCATCAGTCACATGCA AA (SEQ ID NO: 357)	350
34.9 9897 9	DG13S15 22	AGATCTCCAGGGCAGAGG AC(SEQ ID NO: 358)	CCTTCCTCCCTCCTTCTC TC(SEQ ID NO: 359)	355
35.0 7496 2	DG13S15 17	CGTCATTGATCCCAATCAT CT(SEQ ID NO: 360)	GGCTGATAGCCTCCCTTG TA (SEQ ID NO:361)	235
35.0 7496 2	DG13S15 21	GAGAGAGAGCAGCTTGCA TGT(SEQ ID NO:362)	GGCTGATAGCCTCCCTTG TA(SEQ ID NO:363)	172
35.1 2688 2	DG13S36 4	ACCTTTCAAGCTTCCGGTT T(SEQ ID NO: 364)	TTCCATCCGTCCATCTAT CC(SEQ ID NO: 365)	172
35.3 2866 3	DG13S10 36	TTAAAGTCACTTGTCTGTG GTCA(SEQ ID NO: 366)	TTTGTAGGAATCAAGTCA AATAATGTA(SEQ ID NO: 367)	216
35.3 3536 4	DG13S36 7	CAAACATCACACTGGGCAA A(SEQ ID NO: 368)	TGCTTTGGAATCTTTCTT GCT(SEQ ID NO: 369)	301
35.3 7195 7	DG13S19 01	CTGCCAGGATGTCAGCATT (SEQ ID NO: 370)	TCCACACTTTCTCATCAC CTAAA(SEQ ID NO: 371)	440
35.4 2029 5	DG13S10 37	CTTTCGGAAGCTTGAGCCT A(SEQ ID NO: 372)	CCCAAGACCACTGCCATA TT(SEQ ID NO: 373)	269
35.4 2584 1	DG13S18 54	TGACAGGTTTGGGTATATT GGA(SEQ ID NO: 374)	TGCTTAATGTAGTGGCAG CA(SEQ ID NO: 375)	124
35.5 0605 3	DG13S10 38	TCCTGCCTTTGTGAATTCC T(SEQ ID NO: 376)	GTTGAATGAGGTGGGCA TTA(SEQ ID NO: 377)	334
35.5 4721 0	DG13S10 39	CCATTTAATCCTCCAGCCA TT(SEQ ID NO: 378)	GCTCCACCTTGTTACCCT GA(SEQ ID NO: 379)	167
35.6 0925 2	DG13S18 40	ACAACCCTGGAATCTGGAC T(SEQ ID NO: 380)	GAAGGAAAGGAAAGGAA AGAAA(SEQ ID NO: 381)	217
35.6	DG13S36	TGACAAGACTGAAACTTCA	GATGCTTGCTTTGGGAG	257

19286	9	TCAG(SEQ ID NO: 382)	GTA(SEQ ID NO: 383)	
35.627911	D13S305	TTGAGGACCTGTCGTTACG (SEQ ID NO: 384)	TTATAGAGCAGTTAAGGCACA (SEQ ID NO: 385)	394
35.656659	DG13S375	TGAGGGTGGTAAGCCCTTATT(SEQ ID NO: 386)	GGAGTTGTGGCCTCTCTCTCT(SEQ ID NO: 387)	192
35.760368	D13S219	AAGCAAATATGCAAAATTGC(SEQ ID NO: 388)	TCCTTCTGTTTCTTGACTTAACA (SEQ ID NO: 389)	125
35.825852	DG13S378	TGCTAAGAGGGCAGATCTCA(SEQ ID NO: 390)	GGCTCATAGCCAATTTCTCC (SEQ ID NO: 391)	324
35.832127	DG13S32	CGGCATTCTCAATAACCTCAA (SEQ ID NO: 392)	TCTTTGATGAGGATCAATTAGTGG (SEQ ID NO: 393)	214
35.872936	DG13S1549	ACGCACACACACACACACAC (SEQ ID NO: 394)	TGCCTCTGTAATCCTGTGTAGC(SEQ ID NO: 395)	260
35.912321	DG13S1473	GCTCTAAGGTGGGTCCCAATA (SEQ ID NO:396)	GGGAATGACAAGATCAGTTTACC (SEQ ID NO: 397)	163

Table 7.

The selected SNP haplotypes and the corresponding p-values, relative risk (RR), number of patients (#aff), allelic frequency in patients (aff.frq.), carrier frequency in patients (carr.frq.),number of controls (#con), allelic frequency in controls (con.frq.), population attributable risk (PAR). The patients used for this analysis were all unrelated within 4 meioses.

	p-val	RR	#aff	aff.frq.	carr.frq.	#con	con.frq.	PAR	DG00AAFIU	SG13S25	DG00AAJFF	DG00AAHII	DG00AAHID	57	SG13S30	SG13S32	SG13S42	SG13S35
B4	4.8E-05	2.08	903	0.106	0.20	619	0.054	0.11		2		2			2		0	
B5	2.4E-05	2.20	910	0.101	0.19	623	0.049	0.11	3	2		2			2		0	
B6	1.8E-06	2.22	913	0.131	0.24	623	0.063	0.14	3	2	2	2				0		2
A4	5.1E-06	1.81	919	0.159	0.29	623	0.095	0.14		2			3	2		0		
A5	2.6E-06	1.91	920	0.150	0.28	624	0.085	0.14	3	2			3	2		0		

EXAMPLE 2 CORRELATION BETWEEN HAPLOTYPES ASSOCIATED WITH MI, AND RISK OF STROKE

Because stroke is a disease that is closely related to MI (occuring on the basis of atherosclerosis), one SNP haplotype in the FLAP gene (haplotype A4, as shown in Table 7) that confers risk to MI was assessed to determine whether it also conferred risk of stroke. This particular 'at risk' haplotype can be defined by the following 4 SNPs: SG13S25 with allele G, DG00AAHID with allele T, B_SNP_310657 with allele G, and SG13S32 with allele A.

Table 8 shows that the haplotype (A4) increases the risk of having a stroke to a similar extent as it increases the risk of having an MI. The 'at risk' haplotype is carried by 28% of stroke patients and 17% of controls, meaning that the relative risk of having stroke for the carriers of this haplotype is 1.7 (p-value = 5.8 10^{-06}).

Table 8.

		p-val	r	#aff	aff.frq.	#con	con.frq.	info	SG13S6	SG13S25	DG00AAJFF	DG00AAFJT	DG00AAHII	DG00AAHID	SG13S26	B_SNP_310657	SG13S30	SG13S32	SG13S41	SG13S42
MI haplotypes																				
All MI patients																				
	A4	5.3E-07	1.80	1407	0.16	614	0.09	0.82		2				3		2		0		
	B4	1.0E-04	1.87	1388	0.10	612	0.06	0.67		2			2				2			0
Males MI																				
	A4	2.5E-08	2.00	864	0.17	614	0.09	0.82		2				3		2		0		
	B4	1.1E-05	2.12	852	0.11	612	0.06	0.67		2			2				2			0
Females MI																				
	A4	1.9E-02	1.44	543	0.13	614	0.09	0.73		2				3		2		0		
	B4	7.9E-02	1.45	536	0.08	612	0.06	0.60		2			2				2			0
Replication in stroke																				
All stroke patients																				
	A4	5.8E-06	1.73	1238	0.15	614	0.09	0.80		2				3		2		0		
	B4	2.3E-04	1.83	1000	0.10	612	0.06	0.71		2			2				2			0
Males stroke																				
	A4	1.1E-06	1.91	710	0.17	614	0.09	0.79		2				3		2		0		
	B4	3.1E-05	2.11	574	0.11	612	0.06	0.72		2			2				2			0
Females stroke																				
	A4	9.9E-03	1.49	528	0.13	614	0.10	0.74		2				3		2		0		
	B4	6.3E-02	1.47	426	0.08	612	0.06	0.70		2			2				2			0
All stroke excluding MI		8.4E-05	1.65	1054	0.15	614	0.09	0.78		2				3		2		0		
Males stroke excluding MI		6.4E-05	1.78	573	0.16	614	0.09	0.75		2				3		2		0		
Females stroke excluding MI		1.2E-02	1.49	481	0.14	614	0.10	0.72		2				3		2		0		

Cardioembolic stroke	6.6E-04	1.87	248	0.16	614	0.10	0.74	2		3	2		0	
Cardioembolic stroke excluding MI	3.8E-02	1.56	191	0.14	614	0.10	0.70	2		3	2		0	
Large vessel stroke	8.0E-02	1.47	150	0.13	614	0.09	0.83	2		3	2		0	
Large vessel stroke excluding MI	2.9E-01	1.31	114	0.12	614	0.09	0.80	2		3	2		0	
Small vessel stroke	7.2E-04	2.05	166	0.18	614	0.09	0.71	2		3	2		0	
Small vessel stroke excluding MI	1.0E-04	2.31	152	0.20	614	0.10	0.71	2		3	2		0	
Hemorrhagic stroke	4.4E-02	1.73	97	0.15	614	0.09	0.72	2		3	2		0	
Hemorrhagic stroke excluding MI	3.9E-02	1.78	92	0.16	614	0.09	0.71	2		3	2		0	
Unknown cause stroke	1.3E-04	1.88	335	0.16	614	0.09	0.75	2		3	2		0	
Unknown cause stroke excluding MI	6.5E-04	1.82	297	0.16	614	0.09	0.72	2		3	2		0	
MI and stroke together														
All patients														
Best haplo A4	4.1E-07	1.75	2659	0.15	614	0.09	0.82	2		3	2		0	
B4	4.1E-05	1.85	2205	0.10	612	0.06	0.70	2	2			2		0
Males														
A4	1.4E-08	1.93	1437	0.17	614	0.09	0.82	2		3	2		0	
B4	2.0E-06	2.11	1290	0.11	612	0.06	0.70	2	2			2		0
Females														
A4	3.6E-03	1.47	1024	0.13	614	0.09	0.77	2		3	2		0	
B4	2.8E-02	1.48	915	0.08	612	0.06	0.66	2	2			2		0
Patients with both MI and stroke														
A4	6.1E-05	2.10	184	0.18	614	0.09	0.86	2		3	2		0	

EXAMPLE 3 ADDITIONAL CORRELATION BETWEEN FLAP GENE AND MI AND STROKE

A genome wide scan of 296 multiplex Icelandic families with 713 MI patients was performed. The cohort used was a subset of the cohort used in Example 1; in this

cohort, related individuals were assessed. Through the suggestive linkage to a locus on chromosome 13q12-13, the gene encoding the 5-lipoxygenase activating protein (FLAP) was again identified, and a 4-SNP haplotype within the gene was determined to confer a near 2-fold risk of MI and stroke. Male patients showed strongest association to the at-risk haplotype. Independent confirmation of FLAP association to MI was obtained in a British cohort of patients with sporadic MI. These findings support FLAP as the first specific gene isolated that confers substantial risk of the complex traits of MI and stroke.

METHODS

Study population

Patients entering the study were recruited from a registry that includes all MIs that occurred before the age of 75 (over 8,000 patients) in Iceland from 1981 to 2000. This registry is a part of the World Health Organization MONICA Project (The World Health Organization MONICA Project, WHO MONICA Project Principal Investigators,. *J Clin Epidemiol* **41**, 105-14 (1988)). Diagnoses of all patients in the registry follow strict diagnostic rules based on signs, symptoms, electrocardiograms, cardiac enzymes, and necropsy findings.

Genotypes from 713 MI patients and 1741 of their first-degree relatives were used in the linkage analysis. For the microsatellite association study of the MI locus, 802 unrelated MI patients (n=233 females, n=624 males and n= 302 early onset) and 837 population-based controls were used. For the SNP association study in and around the FLAP gene 779 unrelated MI patients were genotyped (n=293 females, n=486 males and n=358 early onset). The control group for the SNP association study was population based and comprised of 628 unrelated males and females in the age range of 30-85 years whose medical history was unknown. The stroke cohort used in this study have previously been described (Gretarsdottir, S. *et al. Nat Genet* **35**, 131-8 (2003); Gretarsdottir, S. *et al., Am J Hum Genet* **70**, 593-603 (2002); Gudmundsson, G. *et al., Am J Hum Genet* **70**, 586-92 (2002)). For the stroke linkage analysis, genotypes from 342 male patients with ischemic stroke or TIA that were

linked to at least one other male patient within and including 6 meioses in 164 families were used. For the association studies 702 patients with all forms of stroke (n=329 females and n=373 males) were analysed. Patients with stroke that also had MI were excluded. Controls used for the stroke association studies were the same as used in the MI SNP association study (n=628).

The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Informed consent was obtained from all study participants. Personal identifiers associated with medical information and blood samples were encrypted with a third party encryption system as previously described (Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K.,. *Eur J Hum Genet* **8**, 739-42 (2000)).

Statistical analysis

A genome-wide scan was performed as previously described (Gretarsdottir, S. *et al. Am J Hum Genet 70*, 593-603 (2002)), using a set of 1000 microsatellite markers. Multipoint, affected-only allele-sharing methods (Kong, A. & Cox, N.J., *Am J Hum Genet 61*, 1179-88 (1997)) were used to assess the evidence for linkage. All results were obtained using the program Allegro (Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, *Nat Genet 25*, 12-3 (2000)) and the deCODE genetic map (Kong, A. *et al., Nat Genet 31*, 241-7 (2002)). The S_{pairs} scoring function (Whittemore, A.S. & Halpern, J., *Biometrics 50*, 118-27 (1994); Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S., *Am J Hum Genet 58*, 1347-63 (1996)) was used, as was the exponential allele-sharing model (Kong, A. & Cox, N.J. *Am J Hum Genet* **61**, 1179-88 (1997)) to generate the relevant 1-df (degree of freedom) statistics. When combining the family scores to obtain an overall score, a weighting scheme was used that is halfway on a log scale between weighting each affected pair equally and weighting each family equally. In the analysis, all genotyped individuals who are not affected are treated as "unknown". Because of concern with small sample behaviour, corresponding P values were usually computed in two different ways for comparison, and the less significant one was reported. The first P value is computed based on large sample theory; $Z_{lr} = \sqrt{(2 \log_e (10) \text{ LOD})}$ and is distributed approximately as a standard

normal distribution under the null hypothesis of no linkage (Kong, A. & Cox, N.J. *Am J Hum Genet* **61**, 1179-88 (1997)). A second P value is computed by comparing the observed LOD score to its complete data sampling distribution under the null hypothesis (Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro,. *Nat Genet* **25**, 12-3 (2000)). When a data set consists of more than a handful of families, these two P values tend to be very similar. The information measure that was used (Nicolae, D. University of Chicago (1999)), and is implemented in Allegro, is closely related to a classical measure of information (Dempster, A., Laird, NM, Rubin, DB., *J R Stat Soc B* **39**, 1-38 (1977) and has a property that is between 0, if the marker genotypes are completely uninformative, and 1, if the genotypes determine the exact amount of allele sharing by descent among the affected relatives.

For single-marker association studies, Fisher's exact test was used to calculate two-sided P values for each allele. All P values were unadjusted for multiple comparisons unless specifically indicated. Allelic rather than carrier frequencies were presented for microsatellites, SNPs and haplotypes. To minimize any bias due to the relatedness of the patients that were recruited as families for the linkage analysis first and second-degree relatives were eliminated from the patient list. For the haplotype analysis, the program NEMO was used (Gretarsdottir, S. *et al.*, *Nat Genet* **35**, 131-8 (2003)), which handles missing genotypes and uncertainty with phase through a likelihood procedure, using the expectation-maximization algorithm as a computational tool to estimate haplotype frequencies. Under the null hypothesis, the affected individuals and controls are assumed to have identical haplotype frequencies. Under the alternative hypotheses, the candidate at-risk haplotype is allowed to have a higher frequency in the affected individuals than in controls, while the ratios of frequencies of all other haplotypes are assumed to be the same in both groups. Likelihoods are maximized separately under both hypotheses, and a corresponding 1-df likelihood ratio statistics used to evaluate statistical significance (*id*). Even though searches were only performed for haplotypes that increase the risk, all reported P values are two-sided unless otherwise stated. To assess the significance of the haplotype association corrected for multiple testing, a randomisation test was carried out using the same genotype data. The cohorts of affected individuals and controls

were randomized, and the analysis was repeated. This procedure was repeated up to 1.000 times and the P value presented is the fraction of replications that produced a P value for a haplotype tested that is lower than or equal to the P value observed using the original patient and control cohorts.

For both single-marker and haplotype analysis, relative risk (RR) and population attributable risk was calculated assuming a multiplicative model (Terwilliger, J.D. & Ott, J. A., *Hum Hered* **42**, 337-46 (1992); Falk, C.T. & Rubinstein, P., *Ann Hum Genet* **51 (Pt 3)**, 227-33 (1987)) in which the risk of the two alleles of haplotypes a person carries multiply. We calculated LD between pairs of SNPs using the standard definition of D′ (Lewontin, R.C., *Genetics* **50**, 757-82 (1964)) and R^2 (Hill, W.G. & Robertson, A., *Genetics* **60**, 615-28 (1968)). Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood, and deviation from linkage equilibrium is evaluated by a likelihood ratio test. When plotting all SNP combinations to elucidate the LD structure in a particular region, D′ was plotted in the upper left corner and the P value in the lower right corner. In the LD plots presented, the markers are plotted equidistantly rather than according to their physical positions.

Identification of DNA polymorphisms.

New polymorphic repeats (i.e., dinucleotide or trinucleotide repeats) were identified with the Sputnik program (http://abajian.net/sputnik/index.html). For microsatellite alleles: the CEPH sample 1347-02 (Centre d'Etudes du Polymorphisme Humain, genomics repository) is used as a reference. The lower allele of each microsatellite in this sample is set at 0 and all other alleles in other samples are numbered according in relation to this reference. Thus allele1 is 1 bp longer than the lower allele in the CEPH sample, allele 2 is 2 bp longer than the lower allele in the CEPH sample, allele 3 is 3 bp longer than the lower allele in the CEPH sample, allele 4 is 4 bp longer than the lower allele in the CEPH sample, allele -1 is 1 bp shorter than the lower allele in the CEPH sample , allele -2 is 2 bp shorter than the lower allele in the CEPH sample, and so on. Single nucleotide polymorphisms in the gene

were detected by PCR sequencing exonic and intronic regions from patients and controls. Public single nucleotide polymorphisms were obtained from the NCBI SNP database. SNPs were genotyped using a method for detecting SNPs with fluorescent polarization template-directed dye-terminator incorporation (SNP-FP-TDI assay) (Chen, X., Zehnbauer, B., Gnirke, A. & Kwok, P.Y., *Proc Natl Acad Sci U S A* 94, 10756-61. (1997)) and TaqMan assays (Applied Biosystems).

British study population

The method of recruitment of 3 separate cohorts of British subjects has been described previously (Steeds, R., Adams, M., Smith, P., Channer, K. & Samani, N.J., *Thromb Haemost* **79**, 980-4 (1998); Brouilette, S., Singh, R.K., Thompson, J.R., Goodall, A.H. & Samani, N.J., *Arterioscler Thromb Vasc Biol* **23**, 842-6 (2003)). In brief, in the first two cohorts a total of 547 patients included those who were admitted to the coronary care units (CCU) of the Leicester Royal Infirmary, Leicester (July 1993–April 1994) and the Royal Hallamshire Hospital, Sheffield (November 1995–March 1997) and satisfied the World Health Organisation criteria for acute MI in terms of symptoms, elevations in cardiac enzymes or electrocardiographic changes (Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. *Circulation* 59, 607-9 (1979)). A total of 530 control subjects were recruited in each hospital from adult visitors to patients with non-cardiovascular disease on general medical, surgical, orthopaedic and obstetric wards to provide subjects likely to be representative of the source population from which the subjects originated. Subjects who reported a history of coronary heart disease were excluded.

In the third cohort, 203 subjects were recruited retrospectively from the registries of 3 coronary care units in Leicester. All had suffered an MI according to WHO criteria before the age of 50 years. At the time of participation, patients were at least 3 months from the acute event. The control cohort comprised 180 subjects with no personal or family history of premature coronary heart disease, matched for age, sex, and current smoking status with the cases. Control subjects were recruited from 3

primary care practices located within the same geographical area. In all cohorts subjects were white of Northern European origin.

RESULTS

Linkage analysis

A genome wide scan was performed in search of MI susceptibility genes using a framework set of 1000 microsatellite markers. The initial linkage analysis included 713 MI patients who fulfilled the WHO MONICA research criteria (The World Health Organization MONICA Project, WHO MONICA Project Principal Investigators,. *J Clin Epidemiol* **41**, 105-14 (1988)) and were clustered in 296 extended families. The linkage analysis was also repeated for early onset, male and female patients separately. Description of the number of patients and families in each analysis are provided in Table 9, and the corresponding allele sharing LOD scores are shown in FIG. 9.

TABLE 9 Number of patients that cluster into families and the corresponding number of families used in the linkage analysis

Phenotype	Number of patients	Number of families	Number of pairs	Genotyped relatives[a]
All MI patients	713	296	863	1741
Males	575	248	724	1385
Females	140	56	108	366
Early onset	194	93	156	739

[a]Genotyped relatives were used to increase the information on IBD sharing

None of these analyses yielded a locus of genome-wide significance. However, the most promising LOD score (LOD = 2.86) was observed on chromosome 13q12 for female MI patients at the peak marker D13S289 (FIG. 9).

This locus also had the most promising LOD score (LOD = 2.03) for patients with early onset MI. After increasing the information on identity-by-descent sharing to over 90% by typing 14 additional microsatellite markers in a 30 centiMorgan (cM) region around D13S289, the LOD score from the female analysis dropped to 2.48 (P value = 0.00036), while the highest LOD score remained at D13S289 (FIG. 10(a)). In addition, in an independent linkage study of male patients with ischemic stroke or transient ischemic attack we observed linkage to the same locus with a LOD score of 1.51 at the same peak marker (FIG. 11), further suggesting that a cardiovascular susceptibility factor might reside at this locus.

Microsatellite association study

The 7.6 Mb region that corresponds to a drop of one in LOD score in the female MI analysis, contains 40 known genes (Table 10).

Table 10 Genes residing within the one LOD drop region of the chromosome 13q12 linkage peak.

LL_Symbol	LL_gene_name
USP12L1	ubiquitin specific protease 12 like 1
RPL21	ribosomal protein L21
GTF3A	general transcription factor IIIA
MTIF3	mitochondrial translational initiation factor 3
PDZRN1	PDZ domain containing ring finger 1
MGC9850	hypothetical protein MGC9850
POLR1D	polymerase (RNA) I polypeptide D, 16kDa
GSH1	GS homeobox 1
IPF1	insulin promoter factor 1, homeodomain transcription factor
CDX2	caudal type homeo box transcription factor 2
FLT3	fms-related tyrosine kinase 3
LOC255967	hypothetical protein LOC255967
FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
C13orf1	chromosome 13 open reading frame 12

2	
LOC283537	hypothetical protein LOC283537
KIAA0774	KIAA0774 protein
SLC7A1	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1
UBL3	ubiquitin-like 3
MGC2599	hypothetical protein MGC2599 similar to katanin p60 subunit A 1 2599
HMGB1	high-mobility group box 1
D13S106E	highly charged protein
ALOX5AP	**arachidonate 5-lipoxygenase-activating protein**
FLJ14834	hypothetical protein FLJ14834
MGC40178	hypothetical protein MGC40178
HSPH1	heat shock 105kDa/110kDa protein 1
B3GTL	beta 3-glycosyltransferase-like
GREAT	similar to G protein coupled receptor affecting testicular descent (H. sapiens)
LOC196549	similar to hypothetical protein FLJ20897
13CDNA73	hypothetical protein CG003
BRCA2	breast cancer 2, early onset
CG018	hypothetical gene CG018
PRO0297	PRO0297 protein
LOC88523	CG016
CG012	hypothetical gene CG012
CG030	hypothetical gene CG030
CG005	hypothetical protein from BCRA2 region
APRIN	androgen-induced proliferation inhibitor
KL	Klotho
STARD13	START domain containing 13
RFC3	replication factor C (activator 1) 3, 38kDa

To determine which gene in this region most likely contributes to MI 120 microsatellite markers were typed within this region, and a case-control association study was performed using 802 unrelated MI patients and 837 population-based controls. The association study was also repeated for each of the three phenotypes that were used in the linkage study, i.e. early onset, male and female MI patients. In addition to testing each marker individually, haplotypes constructed out of those markers for association were also tested. To limit the number of haplotypes tested, only haplotypes that were in excess in the patient cohorts and that spanned less than 300 kb were assessed (see Methods).

As shown in FIG. 10(b), the haplotype that showed association to all MI with the lowest P value (0.00009) covered a region that contains 2 known genes, including the gene encoding arachidonate 5-lipoxygenase-activating protein (FLAP) and a gene with an unknown function called highly charged protein. However, the haplotype association to female MI in this region was less significant (P value =0.005) than for all MI patients and to our surprise, the most significant haplotype association was observed for male MI patients (P value = 0.000002). This male MI haplotype was the only haplotype that remained significant after adjusting for all haplotypes tested.

In view of the association results described above, FLAP was an attractive candidate and therefore efforts were focused on this gene.

Screening for polymorphisms in FLAP and linkage disequilibrium mapping

To determine whether variations within the FLAP gene significantly associate with M,I and to search for causal variations, the FLAP gene was sequenced in 93 patients and 93 controls. The sequenced region covers 60 kb containing the FLAP gene, including the 5 known exons and introns and the 26 kb region 5′ to the first exon and 7 kb region 3′ to the fifth exon. In all, 144 SNPs were identified, of those 96 were excluded from further analysis either because of low minor allele frequency or they were completely correlated with other SNPs and thus redundant. FIG. 10(c) shows the distribution of the 48 SNPs, used for genotyping, relative to exons, introns and the 5′and 3′flanking regions of the FLAP gene. Only one SNP was identified within a coding sequence (exon 2). This SNP did not lead to amino acid substitution.

The locations of these SNPs in the NCBI human genome assembly, build 34, are listed in Table 11.

Table 11: Locations of all genotyped SNPs in NCBI build 34 of the human genome assembly.

SNP name	Build34 start
SG13S381	29083350
SG13S366	29083518
SG13S1	29086224
SG13S2	29087473
SG13S367	29088090
SG13S10	29088473
SG13S3	29089044
SG13S368	29089886
SG13S4	29090997
SG13S5	29091307
SG13S90	29091780
SG13S6	29092536
SG13S371	29093964
SG13S372	29094259
SG13S373	29096688
SG13S375	29096874
SG13S376	29096962
SG13S25	29097553
SG13S377	29101965
SG13S100	29104271
SG13S95	29106329
SG13S191	29107830
SG13S106	29108579
SG13S114	29110096
SG13S121	29112174
SG13S122	29112264
SG13S43	29112455
SG13S192	29116308
SG13S88	29116401
SG13S137	29118118
SG13S86	29118815
SG13S87	29118873
SG13S39	29119740
SG13S26	29122253
SG13S27	29122283

SG13S29	29123643
SG13S89	29124441
SG13S96	29124906
SG13S30	29125840
SG13S97	29129139
SG13S32	29130547
SG13S41	29134045
SG13S42	29135877
SG13S34	29137100
SG13S35	29138117
SG13S181	29138633
SG13S184	29139435
SG13S188	29140805

In addition to the SNPs, a polymorphism consisting of a monopolymer A repeat that has been described in the FLAP promoter region was typed (Koshino, T. *et al.,. Mol Cell Biol Res Commun* **2**, 32-5 (1999)).

The linkage disequilibrium (LD) block structure defined by the 48 SNPs that were selected for further genotyping is shown in FIG. 12. A strong LD was detected across the FLAP region, although it appears that at least one recombination may have occurred dividing the region into two strongly correlated LD blocks.

Haplotype association to MI

To perform a case-control association study the 48 selected SNPs and the monopolymer A repeat marker were genotyped in a set of 779 unrelated MI patients and 628 population-based controls. Each of the 49 markers was tested individually for association to the disease. Three SNPs, one located 3 kb upstream of the first exon and the other two 1 and 3 kb downstream of the first exon, showed nominally significant association to MI (Table 12).

Table 12 SNP allelic association in the MI cohort

Phenotype	Marker	Allele	*P* value	RR	# Pat.	% Pat.	# Ctrl	% Ctrl
All patients	SG13S106	G	0.0044	1.29	681	72.0	530	66.6
	SG13S100	A	0.020	1.29	388	69.6	377	63.9
	SG13S114	T	0.021	1.21	764	70.0	602	65.8
Males	SG13S106	G	0.0037	1.35	422	72.9	530	66.6
	SG13S100	A	0.0099	1.36	292	70.7	377	63.9
	SG13S114	T	0.026	1.24	477	70.4	602	65.8
Early onset	SG13S100	A	0.0440	1.43	99	71.7	377	63.9

Nominally significant SNP association with corresponding number of patients (# Pat.) and controls (#Ctrl). RR refers to relative risk.

However, after adjusting for the number of markers tested, these results were not significant. A search was then conducted for haplotypes that show association to the disease using the same cohorts. For computational reasons, the search was limited to haplotype combinations constructed out of two, three or four SNPs and only haplotypes that were in excess in the patients were tested. The resulting P values were adjusted for all the haplotypes we tested by randomizing the patients and controls (see Methods).

Several haplotypes were found that were significantly associated to the disease with an adjusted P value less that 0.05 (Table 13).

TABLE 13 SNP haplotypes that significantly associate with Icelandic MI patients

SG13S4	SG13S6	SG13S372	SG13S25	SG13S377	SG13S100	SG13S95	SG13S114	SG13S192	SG13S137	SG13S86	SG13S87	SG13S39	SG13S27	SG13S89	SG13S96	SG13S32	SG13S41	SG13S42	SG13S34	SG13S188	*P* value [a]	*P* value [b]	Pat.frq	Ctrl.frq	RR	D' [c]
			G				T							G		A					0,0000023	0,005	0,158	0,095	1,80	1,00
			G				T				A					A					0,0000030	0,006	0,158	0,095	1,78	1,00
			G				T									A			T		0,0000032	0,007	0,157	0,094	1,79	1,00
			G		A						A					A					0,0000046	0,012	0,158	0,083	2,07	0,89
			G			T	T									A					0,0000047	0,012	0,154	0,093	1,78	1,00
			G				T			G						A					0,0000055	0,015	0,147	0,087	1,81	1,00
			G		A											A			T		0,0000061	0,017	0,157	0,083	2,07	0,89
			G		A									G		A					0,0000063	0,017	0,157	0,084	2,04	0,89
			G				T									A					0,0000070	0,021	0,157	0,096	1,76	1,00
			G				T								A	A					0,0000075	0,022	0,149	0,089	1,78	1,00
	G					T	T									A					0,0000083	0,024	0,208	0,139	1,62	0,99
			G		A					G						A					0,0000084	0,026	0,145	0,074	2,14	0,88
			G				T	A								A					0,0000084	0,026	0,139	0,082	1,82	1,00
			G				T						G			A					0,0000091	0,028	0,156	0,096	1,75	1,00
	G						T									A			T		0,0000094	0,028	0,210	0,141	1,61	0,99
	G		G				T									A					0,0000100	0,028	0,156	0,096	1,74	1,00
	G				A											A				A	0,0000101	0,028	0,215	0,133	1,80	0,81
			G		A											A					0,0000105	0,028	0,157	0,084	2,03	0,89
	G				A						A					A					0,0000108	0,029	0,214	0,133	1,78	0,81
			G		A										A	A					0,0000110	0,030	0,146	0,075	2,10	0,88
	G						T				A					A					0,0000112	0,030	0,212	0,144	1,60	1,00
			G		A			A											T		0,0000113	0,030	0,151	0,081	2,03	0,78
			G				T					G				A					0,0000118	0,031	0,156	0,096	1,73	1,00
	G				A											A			T		0,0000126	0,034	0,212	0,131	1,79	0,79
	G						T							G		A					0,0000129	0,035	0,211	0,144	1,59	1,00
			G		A								G			A					0,0000134	0,035	0,156	0,084	2,01	0,89
	G						T									A					0,0000136	0,036	0,211	0,143	1,60	1,00
	G		G		A											A					0,0000137	0,036	0,156	0,085	2,00	0,89
			G		A			A							A						0,0000148	0,037	0,151	0,081	2,01	0,78
			G				T	A											T		0,000015	0,037	0,160	0,099	1,73	0,87

SG13S4	SG13S6	SG13S372	SG13S25	SG13S377	SG13S100	SG13S95	SG13S114	SG13S192	SG13S137	SG13S86	SG13S87	SG13S39	SG13S27	SG13S89	SG13S96	SG13S32	SG13S41	SG13S42	SG13S34	SG13S188	*P* value[a]	*P* value[b]	Pat.frq	Ctrl.frq	RR	D'[c]
																					0					
			G		A			A								A					0,0000150	0,037	0,130	0,066	2,13	0,90
			G				T		C										T		0,0000154	0,039	0,152	0,094	1,73	0,93
			G				T									A		A			0,0000154	0,040	0,155	0,097	1,70	1,00
			G				T		C							A					0,0000157	0,040	0,141	0,085	1,76	1,00
			G	G	A											A					0,0000158	0,040	0,152	0,084	1,94	0,90
	G						T						G			A					0,0000163	0,040	0,210	0,143	1,59	0,99
	G						T			G						A					0,0000166	0,041	0,200	0,134	1,61	0,92
	G				A									G		A					0,0000168	0,042	0,213	0,133	1,76	0,81
			G		A							G				A					0,0000168	0,042	0,156	0,084	2,00	0,89
C	G				A											A					0,0000171	0,042	0,211	0,136	1,70	0,81
	G						T	A								A					0,0000183	0,043	0,192	0,128	1,62	0,85
	G				A											A					0,0000184	0,043	0,212	0,132	1,77	0,81
	G						T										A		T		0,0000193	0,046	0,328	0,251	1,46	0,99
			G				T						G						T		0,0000194	0,046	0,175	0,115	1,64	0,98
	G	G			A											A					0,0000202	0,048	0,210	0,136	1,70	0,81
	G		G		A			A													0,0000209	0,049	0,151	0,082	2,00	0,76

[a] Single test P values. [b] *P* values adjusted for all the SNP haplotypes tested. [c]Measure of correlation with haplotype A4.

The most significant association was observed for a four SNP haplotype spanning 33 kb, including the first four exons of the gene (Fig. 10(c)), with a nominal P value of 0.0000023 and an adjusted P value of 0.005. This haplotype, labelled haplotype A4, has haplotype frequency of 15.8% (carrier frequency 30.3%) in patients versus 9.5% (carrier frequency 17.9%) in controls (Table 14).

Table 14: Association of the A4 haplotype to MI and Stroke

Phenotype (n)	Frq. Pat.	RR	PAR	*P*-value	*P*-value[a]
MI (779)	0.158	1.80	0.135	0.0000023	0.005
Males (486)	0.169	1.95	0.158	0.00000091	ND[b]
Females (293)	0.138	1.53	0.094	0.0098	ND
Early onset (358)	0.138	1.53	0.094	0.0058	ND
Stroke (702)[c]	0.149	1.67	0.116	0.000095	ND
Males (373)	0.156	1.76	0.131	0.00018	ND
Females (329)	0.141	1.55	0.098	0.0074	ND

[a] *P* value adjusted for the number of haplotypes tested. [b]Not done. [c]Excluding known cases of MI.

Shown is the FLAP A4 haplotype and corresponding number of patients (n), haplotype frequency in patients (Frq. pat.), relative risk (RR), population attributed risk (PAR) and *P* values. The A4 haplotype is defined by the following SNPs: SG13S25, SG13S114, SG13S89 and SG13S32 (Table 13). The same controls (n=628) are used for the association analysis in MI and stroke, as well as for the male, female and early onset analysis. The A4 frequency in the control cohort is 0.095.

The relative risk conferred by Haplotype A4 compared to other haplotypes constructed out of the same SNPs, assuming a multiplicative model, was 1.8 and the corresponding population attributable risk (PAR) was 13.5%. As shown in Table 14, Haplotype A4 was observed in higher frequency in male patients (carrier frequency 30.9%) than in female patients (carrier frequency 25.7%). All the other haplotypes that were significantly associated with an adjusted P value less than 0.05, were highly

correlated with Haplotype A4 and should be considered variants of that haplotype (Table 13).

Association of Haplotype A4 to stroke

In view of the linkage observed for stroke in male patients to the FLAP locus and since there is a high degree of co-morbidity among MI and stroke, with most of these cases occurring on the basis of an atherosclerotic disease, it was determined whether Haplotype A4 also shows association to stroke and typed the SNPs defining Haplotype A4 on these patient cohorts. First and second degree relatives and all known cases of MI were removed, and 702 stroke patients were tested for association. The results are also listed in Table 14, above. A significant association of Haplotype A4 to stroke was observed, with a relative risk of 1.67 (P value = 0.000095). In addition, it was determined whether Haplotype A4 was primarily associated with a particular sub-phenotype of stroke, and found that both ischemic and hemorrhagic stroke were significantly associated with Haplotype A4 (Table 15).

Table 15: Association of Haplotype A4 to subgroups of stroke

Phenotype (n)	Pat. Frq.	RR	PAR	P-value
Stroke[a] (702)	0.149	1.67	0.116	0.000095
Ischemic (484)	0.148	1.65	0.113	0.00053
TIA (148)	0.137	1.51	0.090	0.058
Hemorrhagic (68)	0.167	1.91	0.153	0.024

[a]Excluding known cases of MI.

It should be noted that similar to the stronger association of Haplotype A4 to male MI compared to female MI, it also shows stronger association to male stroke (Table 14).

Haplotype association to FLAP in a British cohort

In an independent study, it was determined whether variants in the FLAP gene also have impact on risk of MI in a population outside Iceland. The four SNPs,

defining Haplotype A4, were typed in a cohort of 750 patients from the United Kingdom who had sporadic MI, and in 728 British population controls. The patients and controls come from 3 separate study cohorts recruited in Leicester and Sheffield. No significant differences were found in the frequency of the haplotype between patients and controls (16.9% versus 15.3%, respectively). However, when we typed additional 9 SNPs, distributed across the FLAP gene, in the British cohort and searched for other haplotypes that might be associated with MI, two SNPs showed association to MI with a nominally significant P value (data not shown). Moreover, three and four SNP haplotype combinations increased the risk of MI in the British cohort further and the most significant association was observed for a four SNP haplotype with a nominal P value = 0.00037 (Table 16).

Table 16: Association of the HapB haplotype to British MI patients

Phenotype (n)	Frq. Pat.	RR	PAR	*P*-value	*P*-value[a]
MI (750)	0.075	1.95	0.072	0.00037	0.046
Males (546)	0.075	1.97	0.072	0.00093	ND
Females (204)	0.073	1.90	0.068	0.021	ND

[a]*P* value adjusted for the number of haplotypes tested using 1,000 randomization tests.

Shown are the results for HapB that shows the strongest association in British MI cohort. HapB is defined by the following SNPs: SG13S377, SG13S114, SG13S41 and SG13S35 (that have the following alleles A, A, A and G, respectively. In all three phenotypes shown the same set of n=728 British controls is used and the frequency of HapB in the control cohort is 0.040. Number of patients (n), haplotype frequency in patients (Frq. pat.), relative risk (RR) and population attributed risk (PAR).

This was called haplotype HapB. The haplotype frequency of HapB is 7.5% in the MI patient cohort (carrier frequency 14.4%), compared to 4.0% (carrier frequency 7.8%) in controls, conferring a relative risk of 1.95 (Table 16). This haplotype remained significant after adjusting for all haplotypes tested, using 1000 randomisation steps, with an adjusted P value = 0.046. No other SNP haplotype had an adjusted P value less than 0.05. The two at-risk haplotypes haplotype A4 and

HapB appear to be mutually exclusive with no instance where the same chromosome carries both haplotypes.

DISCUSSION:

These results show that variants of the gene encoding FLAP associate with increased risk of MI and stroke. In the Icelandic cohort, a haplotype that spans the FLAP gene is carried by 30% of all MI patients and almost doubles the risk of MI. These findings were subsequently replicated in an independent cohort of stroke patients. In addition, another haplotype that spans the FLAP gene is associated with MI in a British cohort. Suggestive linkage to chromosome 13q12 was observed with several different phenotypes, including female MI, early onset MI of both sexes, and ischemic stroke or TIA in males. However, surprisingly, the strongest haplotype association was observed to males with MI or stroke. Therefore, there may be other variants or haplotypes within the FLAP gene, or in other genes within the linkage region, that also may confer risk to these cardiovascular phenotypes.

These data also show that the at-risk haplotype of the FLAP gene has increased frequency in all subgroups of stroke, including ischemic, TIA, and hemorrhagic stroke.

Association was not found between Haplotype A4 and MI in a British cohort. However, significant association to MI was found with a different variant over the FLAP gene. The fact that different haplotypes of the gene are found conferring risk to MI in a second population is not surprising. A common disease like MI associates with many different mutations or sequence variations, and the frequencies of these disease associated variants may differ between populations. Furthermore, the same mutations may be seen arising on different haplotypic backgrounds.

All references cited herein are incorporated by reference in their entirety. While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.